Title: In Vivo Characterization of a Vertebrate Ultra-conserved Enhancer

Corresponding Author: Dr. Len Pennacchio,

Corresponding Author's Institution: Lawrence Berkeley National Laboratory

First Author: Francis Poulin

Order of Authors: Francis Poulin; Marcelo A Nobrega; Ingrid Plajzer-Frick; Amy Holt; Veena Afzal; Edward M Rubin; Len Pennacchio

Manuscript Region of Origin:

Abstract: Genomic sequence comparisons between human, mouse and pufferfish (Takifugu rubripes (Fugu)) have revealed a set of extremely conserved noncoding sequences. While this high degree of sequence conservation suggests severe evolutionary constraint and predicts a lack of tolerance to change in order to retain in vivo functionality, such elements have been minimally explored experimentally. In this study, we describe the in-depth characterization of an ancient conserved enhancer, Dc2 located near the dachshund gene, which displays a human-Fugu identity of 84% over 424 basepairs (bp). In addition to this large overall conservation, we find that Dc2 is characterized by the presence of a large block of sequence (144 bp) that is completely identical between human, mouse, chicken, zebrafish and Fugu. Through the testing of reporter vector constructs in transgenic mice, we observed that the 424 bp Dc2 conserved element is necessary and sufficient for brain tissue enhancer activity. In vivo analyses also revealed that the 144 bp 100% conserved sequence is necessary, but not sufficient, to replicate Dc2 enhancer function. However, the introduction of two separate 16 bp insertions into the highly conserved enhancer core did not cause any detectable

modification of its in vivo activity. Our observations indicate that the 144 bp 100% conserved element is tolerant of change at least at the resolution of this transgenic mouse assay and suggest that purifying selection on Dc2 sequence might not be as strong as we predicted or that some unknown property also constrains this highly conserved enhancer sequence.

2800 Mitchell Drive, Walnut Creek, CA 94598   One Cyclotron Road, Berkeley, CA 94720

March 4, 2005

Dr. Richard J Mural
Editor, Genomics

Dear Dr. Mural:

Thank you for your letter of February 15th and for the reviewers' comments on our manuscript entitled "*In Vivo* Characterization of a Vertebrate Ultra-conserved Enhancer", by Poulin et al. We have addressed the reviewers' concerns in a revised manuscript, which we are resubmitting for your consideration.

Our detailed response is as follows:

Reviewer #2

1.   Abstract - last sentence should read: "tolerant of change at least AT the resolution…"

We have changed this typo.

2.   Introduction paragraph four (page 4), first sentence. Statement "which is involved in embryonic development" is too vague. Please state, briefly, the function of DACH.

We have added an introduction description of DACHs known function and reference to its Drosophila homolog.

3.   Introduction paragraph four (page 4). When first introducing Dc2, please define it as a conserved, functional element at DACH (i.e., since the Dc nomenclature was not used previously in the text).

We have made this change**.**

4.   Introduction paragraph four (page 4). Please provide a reference for the statement "known degeneracy of the sequences recognized by transcription factors".

We have added a reference.

5.   Results paragraph one (page 5), first sentence. Please restate the locus (DACH) at which Dc2 resides.

We have made this change.

6.   Results "Refinement of the Minimal Necessary Dc2 Enhancer". Authors state that the 424 bp fragment drives expression pattern in transgenics that is "indistinguishable" from the original WT fragment. Examining the figure reveals that expression looks "stronger" in WT. Please provide a brief discussion in the discussion section (e.g., that tissue expression look similar but not able to tell

Ph: 510-486-7498
Fax: 510-486-4229
Email: LAPennacchio@lbl.gov
www.jgi.doe.gov

if expression levels are affected).

We have modified the corresponding results section to specify that it is the expression pattern that is indistinguishable. In addition, we now mention in the discussion (p. 12) that our assay might not detect differences in expression levels.

7.   Results "The Highly Conserved Minimal Enhancer Core….". Authors state that linkers were designed not to introduce transcription factor binding sites. A TRANSFAC 8.1 search using a 6 bp seed revealed Sp1, retinoic acid receptor (alpha and beta), and AP-2 binding sites. Please provide the Genomatix parameters used and briefly discuss limitations in Discussion section.

The Genomatix parameters are now included in the Methods section (p. 13). In addition, we added a discussion of the limitations of our analysis for transcription factor binding sites presence (p. 11).

8.   Text, general. Consider downplaying the point that the enhancer is "tolerant to change". The authors actually argue (very well) in paragraph two of the discussion, why their approach may be insufficient to fully address this issue.

We have used the term "suggest" to qualify that they enhancer may not be tolerant to change. Ultimately an ultra-conserved knockout mouse will be required to address this issue and experiments are in progress towards this goal.

9.   Text, general. In the results and methods the authors state that they used a cut-off of more than 80 bp at least 70% identical for comparative sequence analyses. Please provide a brief rational for this in the discussion.

We have clarified this point in the Results section (p. 5). We selected Dc2 based on its extensive conservation over a region of 424bp, and the plots in Figure 2 more accurately illustrates the extent of the conservation in comparison with the surrounding sequence.

10.   Materials and Methods. General. More details should be provided, even though they may appear throughout the text. For instance:

a.   Provide primer sequences in first sentence of "cloning" for original enhancer amplification or refer to specific line in Table 1.

We now refer to the primers sequence in Table 1.

b.   Provide bp positions and figure reference for linkers in "cloning".

We now include the actual position of the insertion in reference to Figure 2.

c.   As stated earlier, please provide parameters for Genomatix in "cloning".

We have made this change.

Ph: 510-486-7498
Fax: 510-486-4229
Email: LAPennacchio@lbl.gov
www.jgi.doe.gov

d.   Provide reference for injection protocol in "Generation of transgenic mice".

We have made this change.

11.   Figure 3. Add " ' " to Dc2 after DACH2 on lines 3, 6, and 9. Figure may read nicer if consensus is placed on bottom.

We added the " ' " to Dc2 after DACH2, and we also removed the consensus line.

12.   Figure 4. "N/A" seems inappropriate in the right-most column. This implies that the data was not collected, instead of being negative. Please consider changing to "Not Detected", "No Expression", or some alternative abbreviation with an explanatory footnote.

We changed the abbreviation to ND for "Not Detected".

Reviewer #3

1. It would have been informative if the authors had predicted potential transcription factor binding sites in the 424 bp putative enhancer sequence, and examined the distribution of binding sites in relation to the highly conserved 144 bp element and the regions disrupted by 16-bp insertion.

We agree with the reviewer that such a figure would be informative, and we have added a new Figure 3 to the manuscript. This new figure presents the transcription factors binding sites that are conserved between human and fugu in the 424bp enhancer, as well as the position of the two insertions.

We thank the reviewers for their constructive and thoughtful criticisms. We feel that we have responded satisfactorily to their comments, and that the paper is now acceptable for publication in *Genomics*.

Yours sincerely,

Len A. Pennacchio, Ph.D

Department Head, Vertebrate Genome Analysis, Joint Genome Institute
and
Staff Scientist, Lawrence Berkeley National Laboratory.
Genome Sciences Department
One Cyclotron Road, MS 84-171
Phone: 510-486-7498
Fax: 510-486-4229
E-mail: lapennacchio@lbl.gov

Ph: 510-486-7498
Fax: 510-486-4229
Email: LAPennacchio@lbl.gov
www.jgi.doe.gov

JGI

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

BERKELEY LAB

2800 Mitchell Drive, Walnut Creek, CA 94598   One Cyclotron Road, Berkeley, CA 94720

March 4, 2005

Dr. Richard J Mural
Editor, Genomics

Dear Dr. Mural:

Thank you for your letter of February 15th and for the reviewers' comments on our manuscript entitled "*In Vivo* Characterization of a Vertebrate Ultra-conserved Enhancer", by Poulin et al. We have addressed the reviewers' concerns in a revised manuscript, which we are resubmitting for your consideration.

Our detailed response is as follows:

Reviewer #2

1.   Abstract - last sentence should read: "tolerant of change at least AT the resolution…"

We have changed this typo.

2.   Introduction paragraph four (page 4), first sentence. Statement "which is involved in embryonic development" is too vague. Please state, briefly, the function of DACH.

We have added an introduction description of DACHs known function and reference to its Drosophila homolog.

3.   Introduction paragraph four (page 4). When first introducing Dc2, please define it as a conserved, functional element at DACH (i.e., since the Dc nomenclature was not used previously in the text).

We have made this change**.**

4.   Introduction paragraph four (page 4). Please provide a reference for the statement "known degeneracy of the sequences recognized by transcription factors".

We have added a reference.

5.   Results paragraph one (page 5), first sentence. Please restate the locus (DACH) at which Dc2 resides.

We have made this change.

6.   Results "Refinement of the Minimal Necessary Dc2 Enhancer". Authors state that the 424 bp fragment drives expression pattern in transgenics that is "indistinguishable" from the original WT fragment. Examining the figure reveals that expression looks "stronger" in WT. Please provide a brief discussion in the discussion section (e.g., that tissue expression look similar but not able to tell

Ph: 510-486-7498
Fax: 510-486-4229
Email: LAPennacchio@lbl.gov
www.jgi.doe.gov

if expression levels are affected).

We have modified the corresponding results section to specify that it is the expression pattern that is indistinguishable. In addition, we now mention in the discussion (p. 12) that our assay might not detect differences in expression levels.

7.   Results "The Highly Conserved Minimal Enhancer Core….". Authors state that linkers were designed not to introduce transcription factor binding sites. A TRANSFAC 8.1 search using a 6 bp seed revealed Sp1, retinoic acid receptor (alpha and beta), and AP-2 binding sites. Please provide the Genomatix parameters used and briefly discuss limitations in Discussion section.

The Genomatix parameters are now included in the Methods section (p. 13). In addition, we added a discussion of the limitations of our analysis for transcription factor binding sites presence (p. 11).

8.   Text, general. Consider downplaying the point that the enhancer is "tolerant to change". The authors actually argue (very well) in paragraph two of the discussion, why their approach may be insufficient to fully address this issue.

We have used the term "suggest" to qualify that they enhancer may not be tolerant to change. Ultimately an ultra-conserved knockout mouse will be required to address this issue and experiments are in progress towards this goal.

9.   Text, general. In the results and methods the authors state that they used a cut-off of more than 80 bp at least 70% identical for comparative sequence analyses. Please provide a brief rational for this in the discussion.

We have clarified this point in the Results section (p. 5). We selected Dc2 based on its extensive conservation over a region of 424bp, and the plots in Figure 2 more accurately illustrates the extent of the conservation in comparison with the surrounding sequence.

10.   Materials and Methods. General. More details should be provided, even though they may appear throughout the text. For instance:

a.   Provide primer sequences in first sentence of "cloning" for original enhancer amplification or refer to specific line in Table 1.

We now refer to the primers sequence in Table 1.

b.   Provide bp positions and figure reference for linkers in "cloning".

We now include the actual position of the insertion in reference to Figure 2.

c.   As stated earlier, please provide parameters for Genomatix in "cloning".

We have made this change.

Ph: 510-486-7498
Fax: 510-486-4229
Email: LAPennacchio@lbl.gov
www.jgi.doe.gov

d.   Provide reference for injection protocol in "Generation of transgenic mice".

We have made this change.

11.   Figure 3. Add " ' " to Dc2 after DACH2 on lines 3, 6, and 9. Figure may read nicer if consensus is placed on bottom.

We added the " ' " to Dc2 after DACH2, and we also removed the consensus line.

12.   Figure 4. "N/A" seems inappropriate in the right-most column. This implies that the data was not collected, instead of being negative. Please consider changing to "Not Detected", "No Expression", or some alternative abbreviation with an explanatory footnote.

We changed the abbreviation to ND for "Not Detected".

Reviewer #3

1. It would have been informative if the authors had predicted potential transcription factor binding sites in the 424 bp putative enhancer sequence, and examined the distribution of binding sites in relation to the highly conserved 144 bp element and the regions disrupted by 16-bp insertion.

We agree with the reviewer that such a figure would be informative, and we have added a new Figure 3 to the manuscript. This new figure presents the transcription factors binding sites that are conserved between human and fugu in the 424bp enhancer, as well as the position of the two insertions.

We thank the reviewers for their constructive and thoughtful criticisms. We feel that we have responded satisfactorily to their comments, and that the paper is now acceptable for publication in *Genomics*.

Yours sincerely,

Len A. Pennacchio, Ph.D

Department Head, Vertebrate Genome Analysis, Joint Genome Institute
and
Staff Scientist, Lawrence Berkeley National Laboratory.
Genome Sciences Department
One Cyclotron Road, MS 84-171
Phone: 510-486-7498
Fax: 510-486-4229
E-mail: lapennacchio@lbl.gov

Ph: 510-486-7498
Fax: 510-486-4229
Email: LAPennacchio@lbl.gov
www.jgi.doe.gov

In Vivo **Characterization of a Vertebrate Ultra-conserved Enhancer**

Francis Poulin*****, Marcelo A. Nobrega, Ingrid Plajzer-Frick, Amy Holt, Veena Afzal, Edward M. Rubin, and Len A. Pennacchio

DOE Joint Genome Institute

Walnut Creek, CA 94598

and

Genomics Division

Lawrence Berkeley National Laboratory

Berkeley, CA 94720

[*]Present address:　Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720-3140

To whom correspondence should be addressed:

Len A. Pennacchio, Genomics Division, One Cyclotron Road, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Email: LAPennacchio@lbl.gov, Phone: (510) 486-7498, Fax: (510) 486-4229

1

Abstract

Genomic sequence comparisons between human, mouse and pufferfish (*Takifugu rubripes (Fugu)*) have revealed a set of extremely conserved noncoding sequences. While this high degree of sequence conservation suggests severe evolutionary constraint and predicts a lack of tolerance to change in order to retain *in vivo* functionality, such elements have been minimally explored experimentally. In this study, we describe the in-depth characterization of an ancient conserved enhancer, Dc2 located near the dachshund gene, which displays a human-*Fugu* identity of 84% over 424 basepairs (bp). In addition to this large overall conservation, we find that Dc2 is characterized by the presence of a large block of sequence (144 bp) that is completely identical between human, mouse, chicken, zebrafish and *Fugu*. Through the testing of reporter vector constructs in transgenic mice, we observed that the 424 bp Dc2 conserved element is necessary and sufficient for brain tissue enhancer activity. *In vivo* analyses also revealed that the 144 bp 100% conserved sequence is necessary, but not sufficient, to replicate Dc2 enhancer function. However, the introduction of two separate 16 bp insertions into the highly conserved enhancer core did not cause any detectable modification of its *in vivo* activity. Our observations indicate that the 144 bp 100% conserved element is tolerant of change at least at the resolution of this transgenic mouse assay and suggest that purifying selection on Dc2 sequence might not be as strong as we predicted or that some unknown property also constrains this highly conserved enhancer sequence.

2

Introduction

Vertebrate comparative genomics has proven to be an effective approach to uncover functional elements in the human genome. This has been accomplished through comparison of the human genome with the genomes of a wide-range of species from primates to fish [1-4]. The underlying success of this strategy is based on comparing sufficiently divergent genomes to distinguish neutral versus functionally constrained sequence elements [5, 6]. At one extreme, comparison of the human genome to the genome of a teleost fish, *Takifugu rubripes* (*Fugu*), has identified a set of anciently conserved coding and noncoding sequences [7, 8].

Humans and *Fugu* last shared a common ancestor approximately 400 million years ago [9] and while the *Fugu* genome is one of the smallest known in vertebrates (365 Mb), its gene repertoire is similar to that of humans [10]. These characteristics led to the original proposal to sequence the *Fugu* genome in order to assist with the annotation of human genes based on comparative genomics [10]. Furthermore, the availability of *Fugu* genomic sequence also revealed that human-*Fugu* conserved noncoding sequences can be used to delineate gene regulatory sequences [1, 11-15].

Many of the ancient elements conserved between human and *Fugu* overlap with recently described ultra-conserved human/rodent DNA elements [16]. Through the use of human/mouse/rat whole-genome comparisons, ultra-conserved elements were defined as sequences that are at least 200 bp and are absolutely conserved in these three species [16]. Intriguingly, the non-exonic ultra- and human-*Fugu* conserved elements do not show a random

3

distribution in the human genome, but tend to cluster near genes encoding transcription factors that are involved in developmental processes [16, 17].

One example of clustered human-*Fugu* conserved noncoding elements are those surrounding *DACH*, a homolog of the *Drosophila dachshund* gene [18]. In both vertebrates and invertebrates *dachshund* displays a complex temporal and spatial pattern of expression, and the gene product is critical in the development of the central nervous system, sensory organs, and limbs [19-22, 23]. *In vivo* analysis of nine DNA elements conserved from human to *Fugu* demonstrated that seven displayed enhancer activities recapitulating some aspects of *DACH* expression [15]. The most surprising characteristic of these enhancers is their extremely high degree of conservation between humans and fish. For instance Dc2, a conserved regulatory element of human *DACH* that is known to drive expression in hindbrain, forebrain and the retina, is 84% identical over 424 bp to its *Fugu* ortholog. As impressive is the high degree of conservation within sub-regions of this element, with portions displaying greater than 99% identity over 270 bp among human, mouse and rat and greater than 98% identity over 219 bp between human and *Fugu*. This is especially striking considering the known degeneracy of the sequences recognized by transcription factors [24], and this observation of extreme conservation suggests that other limitations may prevent these enhancers from changing over evolutionary time. To better characterize the nature of the sequence constraints in highly conserved vertebrate enhancers, we manipulated a single known element, the human Dc2 enhancer, *in vivo* through a series of reporter constructs tested in transgenic mouse assays.

4

Results

In this study, we characterized a single human-*Fugu* enhancer of the *DACH* locus (Dc2) through (1) comparative genomic analyses to define "phylogenetic footprints", (2) deletion constructs in transgenic mice to experimentally define the minimal sequence necessary and sufficient for *in vivo* activity, and (3) targeted mutagenesis to assess whether the enhancer is tolerant of insertional disruption events.

Comparative Genomics Delineates Conserved Modules

The Dc2 enhancer was previously identified by comparing the sequences of the orthologous human and *Fugu DACH* genes [15]. Briefly, a human DNA fragment of 2,086 bp encompassing the sequence conserved in *Fugu* was tested in transgenic mice and demonstrated to enhance the activity of a minimal heat shock promoter (HSP68) fused to β-galactosidase (LacZ) [25]. The Dc2 enhancer was found to drive the expression of LacZ in the hindbrain, forebrain and retina of the developing mouse embryo [15]. In order to better define the sequence elements required for the enhancer function of Dc2, we performed comparative analysis of the functional element in multiple vertebrate species (Fig. 1A).

Comparison of the human, mouse and rat sequences revealed several discrete elements of conservation throughout the enhancer when using an 80bp and 70% identity cut-off, slightly less stringent than the classically defined 100bp and 70% identity previously used in the identification of functional mammalian regulatory elements [3] (Fig. 1A; 4 human-mouse elements). Comparison of the human and chicken sequences shows a decrease in the number of conserved

5

elements, but extensive conservation is still readily observed (Fig. 1A; two human-chicken elements). This is in contrast to the alignment of the human sequence with that of the more distant vertebrates, which highlights a single region of 424 bp that is more than 84% conserved in all the species analyzed (Fig. 1A, Human Dc2 to frog, zebrafish or *Fugu*). This region is 96% identical to the mouse Dc2, and displays 84% identity between the human and *Fugu* Dc2 (Fig. 1A).

Closer inspection of the alignment of this 424 bp region reveals that the largest uninterrupted block of perfect identity between human-mouse is 195 bp and that between human-*Fugu* is 144 bp in length (Fig. 2). This block of perfect identity can be expanded to a remarkable 270 bp with a single mismatch between human and mouse Dc2 (Fig. 2, human nucleotides 75 to 345). Similarly, the human-*Fugu* conservation is 98% for a block of 219 bp from human nucleotides 90 to 308 (Fig. 2). This large block is flanked by shorter regions where the similarity between species is lower (Fig. 2), suggesting that the enhancer is composed of a core sequence that is highly constrained, and satellite sequences that can evolve more rapidly. Analysis of the 424 bp conserved region with rVista 2.0 [26] predicts 111 and 114 putative transcription factor binding sites in the human and *Fugu* sequence, respectively. Of these sites, 72 are conserved in both species (Fig. 3), with 44 sites landing in the 144bp/100% conserved core (61%; Fig. 3, green box).

Comparison of Dc2 to the human genome also uncovered a single additional region of sequence similarity (which we will refer to as Dc2') on chromosome X. Upon detailed further examination of the flanking sequence, we found that Dc2' lies within *DACH2*, a known paralog of *DACH* that arose from an ancient genomic duplication event predating the divergence of the mammalian and

6

fish lineages (Fig. 4). Similar to Dc2, Dc2' is located in the first intron of *DACH2*, with both displaying similar enhancer properties in transgenic mice, further supporting their common origin (data not shown). Alignment of the human Dc2 to the paralogous Dc2' sequence from multiple vertebrates shows a more limited region of similarity between the two elements (Fig. 1B). The human Dc2' contains the most conserved portion of the *DACH* Dc2 enhancer (Fig. 2; 144 bp, Human-*Fugu* Dc2 100% identity), but overall has a lower degree of sequence conservation (Fig. 4; 144 bp, Human Dc2-Dc2' 86% identity).

Refinement of the Minimal Necessary Dc2 Enhancer

Using the results from our comparative analysis as a guide, we designed several Dc2 constructs to test for enhancer function in transgenic mouse embryos (Fig. 5). As previously demonstrated [15], we found that the wild-type construct displayed enhancer activity in all transgenic lines tested (Fig. 5B, WT), with specific staining in the forebrain, hindbrain and retina (Fig. 5B, WT). To define the minimum fragment necessary for this activity, we tested a construct consisting of only the core 424 bp of sequence that are conserved in the *DACH* Dc2 orthologues from human to *Fugu* (Fig. 5B, 424 bp/84%). We found this 424 bp fragment consistently drove expression in transgenic embryos in a pattern indistinguishable from the parent 2,086 bp construct. Conversely, we found that deletion of this 424 bp sequence from the parent (WT) construct completely abolished its activity in our transgenic assay (Fig. 5B, Δ424bp). This finding suggests the small 424 bp element is able to carry out all the observed enhancer activities found in the full construct.

To further explore the minimal sequence necessary for enhancer activity, we reduced the size of the conserved 424 bp fragment to the 144 bp Dc2 region 100% conserved between human and *Fugu* (Fig. 2, dashed line), a segment that also shows similarity to the paralogous human Dc2' enhancer (Fig. 4). We found that while this 144 bp sequence does not enhance transcription in transgenic mice (Fig. 5B, 144bp/100%), its removal from the WT parental Dc2 construct completely abolished Dc2 enhancer activity (Fig. 5B, Δ144bp). Taken together, these results demonstrate that, although human-mouse comparison shows several regions of sequence conservation (Fig. 1A), the necessary and sufficient portion of the element is contained within the region shared between human and *Fugu*. Moreover, the region of the element that is shared between Dc2 and Dc2' is necessary, but not sufficient, for the function of the enhancer.

The Highly Conserved Minimal Enhancer Core is Tolerant of Insertion Mutations

Having defined the 144 bp core element as essential for enhancer function and 100% conserved between human and *Fugu*, we sought to determine its tolerance to sequence change. The large size of the minimal Dc2 enhancer and its extreme conservation raised several questions regarding the types of constraints acting on its sequence. To address whether the internal organization of the enhancer could limit its variation, we introduced DNA linkers into the phylogenetically conserved core of the enhancer (Fig. 2, arrows). These linkers were 16 bp long, representing an approximately one and a half DNA helix turn, and were designed to not introduce new predicted transcription factor binding sites into the enhancer (see Methods). We chose the location of the insertions to disrupt the most constrained region of the enhancer (Fig. 2, arrows), with one linker (insert 1) located in the segment that is absolutely conserved and has homology to *DACH2* Dc2'

8

(Fig. 2, dashed line).   Surprisingly, in studying the various transgenic lines containing these reporter constructs, we found that the two different linker insertions did not affect the WT reporter construct activity in the mouse embryos (Fig. 5B, Insert 1 and Insert 2), and the pattern of expression of these mutant constructs was not distinguishable from that of the WT construct.

9

Discussion

Comparative genomics is continuing to help localize conserved noncoding sequences with gene regulatory activity, and distant evolutionary comparisons between mammals and teleosts have proven especially efficient [1, 13, 15, 27]. In this study we characterized a gene enhancer (Dc2) identified through human-*Fugu* comparative genomics, and refined the sequences necessary and sufficient for its function, as well as assessed the impact of insertional mutations on its gene regulatory activity.

The Dc2 enhancer is characterized by the presence of a large block of sequence (144 bp) that is entirely identical between human and multiple distantly related species (mouse, rat, chicken, frog, zebrafish, and *Fugu*). This observation is similar to recent reports for mammalian ultra-conserved sequences [16], in spite of the fact that the Dc2 enhancer falls just short of the requirements of the ultra-conserved set (>200 bp, 100% identity). Nonetheless, a similarity in human-*Fugu* and ultra-conserved noncoding sequences is their enormous degree of sequence conservation, extending hundreds of basepairs with minimal substitutions over hundreds of million years of evolution. Regardless of the precise definition of extreme sequence conservation, the question is raised of why such striking constraint? One possible explanation is their enrichment near developmental genes, suggesting they are involved in the tight regulatory control of genes directing the basic vertebrate body plan. But why would gene regulatory sequences require hundreds of basepairs of sequence perfectly conserved over such long time periods when most transcription factors are capable of recognizing short (6-12 bp) degenerate sites? It may be due to a large number of overlapping transcription factor binding sites with

10

highly rigid binding requirements as well as severe constraint on the spacing of putative modules within the enhancer.

Another powerful aspect of vertebrate comparative genomics and gene regulatory sequence characterization is the occurrence of gene paralogs resulting from genomic duplications over the course of evolution. This is readily apparent with the human *DACH* gene and its Dc2 enhancer sharing similarity with the *DACH2* paralog and the adjacent Dc2'. This duplication event is very ancient, predating the last common ancestor of human-*Fugu*, and both loci have strongly resisted sequence change in these functional sequences. Comparison between these paralogs revealed a small and highly conserved core (144 bp) that is necessary to preserve the enhancer function, but is not sufficient to recapitulate it. This suggests that while it is critical to conserve the sequence of certain modules for the enhancer to function, some flanking sequences are more flexible to change. It is known that stabilizing selection can maintain an enhancer function in different species even when sequence conservation is limited because of turnover in transcription factor binding sites [28]. However, the fact that the flanking sequences are flexible in the Dc2 paralogs is contrasted by their extreme conservation in each of the individual orthologs. It therefore appears that the sequence of each of the paralogous elements changed rapidly after the duplication event, and that each eventually became fixed independently [16, 29].

While enormous sequence conservation of human-*Fugu* and ultra-conserved elements alone predicts that any change in such an element would destroy its activity, our analyses indicate that the *DACH* Dc2 enhancer is much more flexible than anticipated. Two separate insertions in the highly conserved core of this enhancer did not cause any detectable modification in its activity *in vivo*. This could be due to the limitations of our assay, which occur at a single time point and

11

lack sensitivity to detect small quantitative changes, or to the presence of another unidentified function within this conserved module. However, the observation that the complexity of the Dc2 expression pattern is not affected by insertions indicates that many of its general activities are still retained. While negative-selection on the Dc2 sequence could not be assessed in these *in vivo* studies, its enormous conservation more likely suggests additional unknown biological properties also constrains Dc2.

12

Materials and Methods

Cloning

Dc2 enhancer constructs were PCR amplified from human genomic DNA (BD Biosciences) and directionally cloned into the pENTR/D-TOPO vector (Invitrogen). The wild-type (WT), 424, and 144 Dc2 constructs were PCR amplified with the corresponding Forward and Reverse primers described in Table 1. Constructs containing site-specific mutations were generated by overlap extension PCR [30], in which the mutagenic primers were used in combination with WT primers of opposite orientation (Table 1). All constructs were sequenced and transferred to the Gateway-HSP68-LacZ vector using the LR recombination reaction (Invitrogen). The elements were cloned in the same orientation relative to the HSP68 promoter as they are to the endogenous *DACH* promoter. A 16bp linker (5'-GCTGCCCGCGCAGTAC) was inserted at two locations (nucleotides 128 and 236 of human *DACH* Dc2, Figure 2) in the wild-type human Dc2 enhancer to test for disruption of Dc2 function. The linker was scanned with the MatInspector software (Genomatix, Matrix Family Library Version 4.2) to ensure the absence of putative transcription-factor binding sites. Vertebrates matrices were used with the following parameters: core similarity: 0.75, matrix silmilarity: optimized.

Generation of transgenic mice

13

Plasmid DNA was purified using the EndoFree plasmid maxi kit (Qiagen). 100 μg of plasmid was linearized with XhoI, followed by purification on Micropure EZ columns and Montage PCR filter units (Millipore). The purified DNA was dialyzed for 24h against injection buffer (10 mM Tris, pH 7.5; 0.1 mM EDTA), and its concentration determined fluorometrically and by agarose gel electrophoresis. The DNA was diluted to a concentration of 1.5 to 2 ng/μl and used for pronuclear injections of FVB embryos [31], in accordance with protocols approved by the Lawrence Berkeley National Laboratory.

Embryo staining

Embryos were harvested at 12.5 dpc and dissected in cold 100mM phosphate buffer pH 7.3, followed by 30min of incubation with 4% paraformaldehyde at 4°C. The embryos' heads were punctured with a 27G needle to facilitate the penetration of the staining solution and washed three times for 30min with wash buffer (2mM $MgCl_2$; 0.01% deoxycholate; 0.02% NP-40; 100mM phosphate buffer, pH 7.3). Embryos were stained for 24h at room temperature with freshly made staining solution (0.8mg/ml X-gal; 4mM potassium ferrocyanide; 4mM potassium ferricyanide; 20mM Tris, pH 7.5 in wash buffer). Stained embryos were rinsed 3 times in 100mM phosphate buffer pH 7.3, and post-fixed in 4% paraformaldehyde. Yolk sacs were carefully dissected from embryos and DNA was prepared by boiling the tissue for 20 min in 75μl of solution 1 (25mM NaOH; 0.2mM EDTA), followed by neutralization with 75μl of solution 2 (40mM Tris-HCl). Yolk sac DNA was screened by PCR with LacZ primers (LacZ-fwd 5'-TTTCCATGTTGCCACTCGC; LacZ-Rev 5'-AACGGCTTGCCGTTCAGCA).

Acknowledgments

Figure legends

Figure 1. Sequence comparison of Dc2 enhancer in multiple species. mVista alignment (http://gsd.lbl.gov/vista/) between the Human *DACH* Dc2 enhancer and (A) orthologous Dc2 sequences from the *DACH* gene, or (B) paralogous Dc2' sequences from the *DACH2* gene of the indicated species. Alignments were performed with an 80 bp window size and a 70% identity threshold.

Figure 2. Sequence alignment of the evolutionarily conserved region from the *DACH* Dc2 enhancer. Dashed lines above the sequence indicate the 144 bp region that is 100% conserved between human and *Fugu* Dc2. Arrowheads indicate the position of insertions in the mutated version of the enhancer. Alignment was preformed using ClustalW.

Figure 3. Conserved transcription factors binding sites in the *DACH* Dc2 enhancer. The 424 bp human and *Fugu* Dc2 enhancers were analyzed for the presence of conserved vertebrates transcription factors binding sites using rVista 2.0 (http://rvista.dcode.org), with a matrix similarity cut-off of 0.9. Alignment between human and Fugu *DACH* Dc2 (424 bp) is depicted as blocks ranging from 50% to 100% conservation. Putative transcription factor binding site positions (identified on the left) are indicated by colored boxes above the alignment. The position of the 144 bp/100% conserved Dc2 core is highlighted in green. Arrowheads indicate the position of insertions in the mutated version of the enhancer.

Figure 4. Local alignment of the human *DACH* Dc2 and *DACH2* Dc2' enhancers. The most conserved portion of the human *DACH* Dc2 enhancer was aligned with the paralogous sequence

16

from *DACH2* Dc2'. Numbering of *DACH* Dc2 nucleotides is the same as in Figure 2. Alignment was performed using ClustalW.

Figure 5. Functional analysis of the human *DACH* Dc2 enhancer. (A) mVista alignment between the Dc2 enhancer from the human and the mouse (H/M), or the *Fugu* (H/F) *DACH* gene. Nucleotide positions are indicated for the human wild-type Dc2 sequence. (B) The indicated DNA fragments from human Dc2 were assayed for *in vivo* enhancer activity on the minimal HSP68 promoter driving LacZ expression. The wild-type (WT) enhancer corresponds to the DNA fragment tested by Nobrega et al. [15]. A 424 bp fragment (nt 318 to 741 of WT) that is 84% identical between human and *Fugu* was tested by itself (424bp/84%), or the corresponding sequence was deleted from the WT fragment (Δ424bp). A 144 bp fragment (nt 404 to 547 of WT) that is 100% identical in human and *Fugu* was tested by itself (144bp/100%), or the corresponding sequence was deleted from the WT fragment (Δ144bp). The WT Dc2 enhancer was modified to insert a 16 bp linker at two different locations in the human-*Fugu* conserved region (Insertion 1 and Insertion 2). The number of mouse embryos displaying the Dc2 expression pattern is reported over the total number of transgenic embryos (Pattern/Tg). ND: Not detected.

17

References

[1] N. Ghanem, et al., Regulatory roles of conserved intergenic domains in vertebrate Dlx bigene clusters, Genome Res. 13 (2003) 533-543.

[2] D. Boffelli, et al., Phylogenetic shadowing of primate sequences to find functional regions of the human genome, Science 299 (2003) 1391-1394.

[3] L.A. Pennacchio and E.M. Rubin, Genomic strategies to identify mammalian regulatory sequences, Nat. Rev. Genet. 2 (2001) 100-109.

[4] J.W. Thomas, et al., Comparative analyses of multi-species sequences from targeted genomic regions, Nature 424 (2003) 788-793.

[5] E.H. Margulies, M. Blanchette, D. Haussler, and E.D. Green, Identification and characterization of multi-species conserved sequences, Genome Res. 13 (2003) 2507-2518.

[6] L. Elnitski, et al., Distinguishing regulatory DNA from neutral sites, Genome Res. 13 (2003) 64-72.

[7] S. Aparicio, et al., Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes, Science 297 (2002) 1301-1310.

[8] S. Aparicio, et al., Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, Fugu rubripes, Proc. Natl Acad. Sci. U.S.A. 92 (1995) 1684-1688.

[9] B. Venkatesh, P. Gilligan, and S. Brenner, Fugu: a compact vertebrate reference genome, FEBS Lett. 476 (2000) 3-7.

[10] S. Brenner, et al., Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome, Nature 366 (1993) 265-268.

[11] D.H. Rowitch, et al., Identification of an evolutionarily conserved 110 base-pair cis-acting regulatory sequence that governs Wnt-1 expression in the murine neural plate, Development 125 (1998) 2735-2746.

[12] H. Marshall, et al., A conserved retinoic acid response element required for early expression of the homeobox gene Hoxb-1, Nature 370 (1994) 567-571.

[13] C. Kimura-Yoshida, et al., Characterization of the pufferfish Otx2 cis-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification, Development 131 (2004) 57-71.

[14] H. Popperl, et al., Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx, Cell 81 (1995) 1031-1042.

18

[15] M.A. Nobrega, I. Ovcharenko, V. Afzal, and E.M. Rubin, Scanning human gene deserts for long-range enhancers, Science 302 (2003) 413.

[16] G. Bejerano, et al., Ultraconserved elements in the human genome, Science 304 (2004) 1321-1325.

[17] D. Boffelli, M.A. Nobrega, and E.M. Rubin, Comparative genomics at the vertebrate extremes, Nat. Rev. Genet. 5 (2004) 456-465.

[18] X. Li, et al., Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis, Nature 426 (2003) 247-254.

[19] S.R. Martini, G. Roman, S. Meuser, G. Mardon, and R.L. Davis, The retinal determination gene, dachshund, is required for mushroom body cell differentiation, Development 127 (2000) 2663-2672.

[20] G. Mardon, N.M. Solomon, and G.M. Rubin, dachshund encodes a nuclear protein required for normal eye and leg development in Drosophila, Development 120 (1994) 3473-3486.

[21] M. Kurusu, et al., Genetic control of development of the mushroom bodies, the associative learning centers in the Drosophila brain, by the eyeless, twin of eyeless, and Dachshund genes, Proc. Natl Acad. Sci. U.S.A. 97 (2000) 2140-2144.

[22] R.J. Davis, W. Shen, T.A. Heanue, and G. Mardon, Mouse Dach, a homologue of Drosophila dachshund, is expressed in the developing retina, brain and limbs, Dev Genes Evol 209 (1999) 526-536.

[23] R.J. Davis, et al., Dach1 mutant mice bear no gross abnormalities in eye, limb, and brain development and exhibit postnatal lethality, Mol Cell Biol 21 (2001) 1484-1490.

[24] G.D. Stormo, DNA binding sites: representation and discovery, Bioinformatics 16 (2000) 16-23.

[25] R. Kothary, et al., Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice, Development 105 (1989) 707-714.

[26] G.G. Loots and I. Ovcharenko, rVISTA 2.0: evolutionary analysis of transcription factor binding sites, Nucleic Acids Res 32 (2004) W217-221.

[27] F. Spitz, F. Gonzalez, and D. Duboule, A global control region defines a chromosomal regulatory landscape containing the HoxD cluster, Cell 113 (2003) 405-417.

[28] M.Z. Ludwig, C. Bergman, N.H. Patel, and M. Kreitman, Evidence for stabilizing selection in a eukaryotic enhancer element, Nature 403 (2000) 564-567.

[29] S. Ohno, Evolution by gene duplication. Springer-Verlag, Berlin, 1970.

19

[30] J. Sambrook and D.W. Russell, Molecular Cloning: a laboratory manual. CSHL Press, Cold Spring Harbor, NY, 2001.

[31] A. Nagy, M. Gertsenstein, K. Vintersten, and R. Behringer, Manipulating the Mouse Embryo: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2002.

Table 1

Table 1 – Primers used to generate Dc2 constructs

| Construct	Forward primer	Reverse primer
WT	5'-GCAATTTTGAAAAAGAAAACAATGG	5'-TAGACAGCTCATGCTGAGAAAACTG
424	5'-AATTCTTTGCCTGATTTTC	5'-TTTTGGTGATGAAGACAG
144	5'-TCAGGGTGCCTTTGAG	5'-GCTCCTTTCATACTTG
Δ424	5'-CTTATTATTAAAATATAGGCTGTCTTCCAGTCTTTGAATAC	5'-GTATTCAAAGACTGGAAGACAGCCTATATTTTAATAATAAG
Δ144	5'-GCCTAAAAAAATCTACTACACATTTCCCTTGGAGCTGC	5'-GCAGCTCCAAGGGAAATGTGTAGTAGATTTTTTTAGGC
Insert 1	5'-GCTGAACGATGTGCATATTCATTAAGGCTCACATA	5'-ATGCACATCGTTCAGCGCAGGATTAGATTTTAATTA
Insert 2	5'-GCTGAACGATGTGCATCCCTTGGAGCTGCCTGC	5'-ATGCACATCGTTCAGCAAATGTGCTCCTTTCATAC

Figure 1
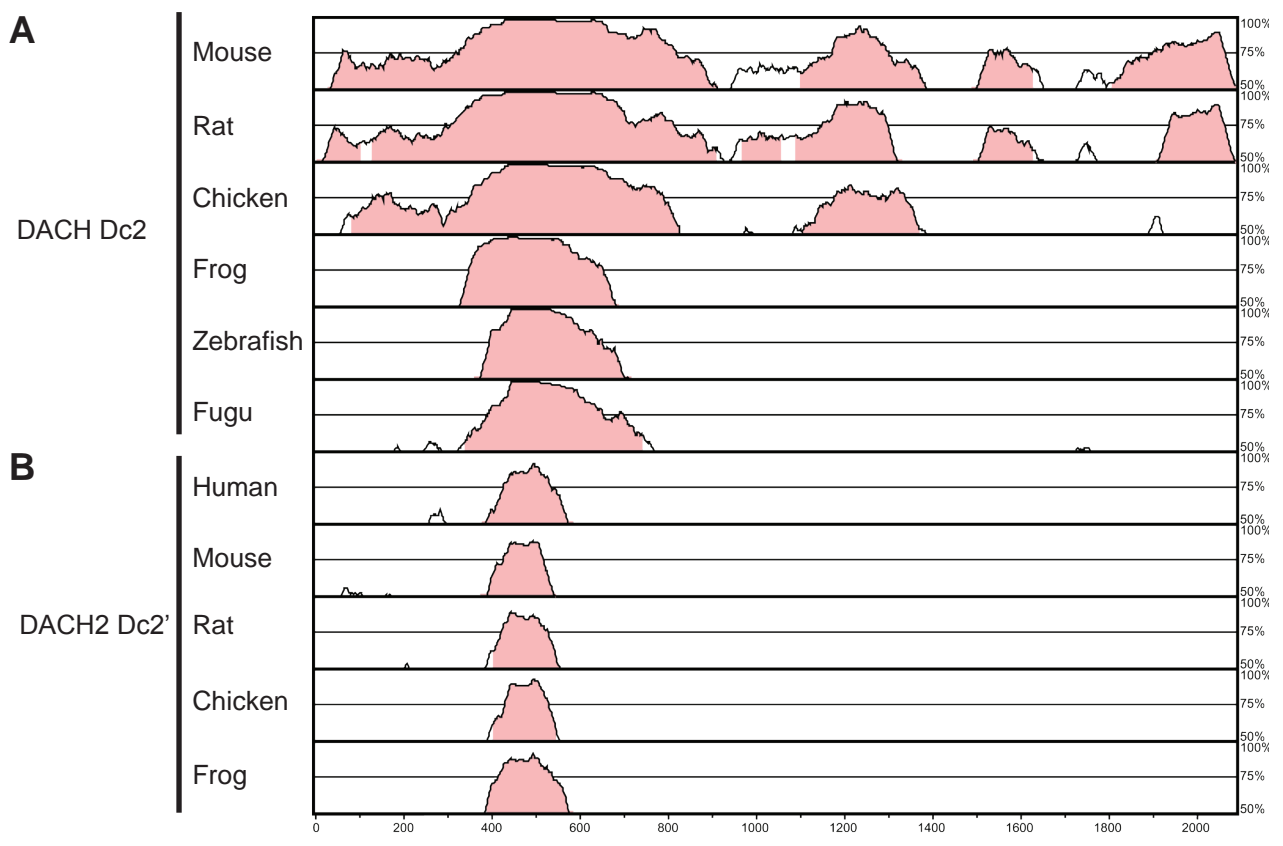
Figure 1 - Poulin et al.

Figure 2

```
Human       1 AATTCTTTGCCTGATTTTCCGCTTTTGTAAAATGCATTGCTTCAGAAACAAGTCTTCAAA
Mouse       1 AGTTCTTTGCCTGCTTTTT-GCTTTTGTAAAATGCATTGCAACAGAAACAAGTCTTCAAA
Rat         1 AGTTCTTTGCCTGCTTTTTTGCTTTTGTAAAATGCATTGCTTCAGAAACAAGTCTTCAAA
Chicken     1 AGTTCTTTGCCTGATTTTCTTCTTTCATAAAATGCACTGCATCTGAAACAAGTCTTCAAA
Xenopus     1 GCTGCTGCGGCTGATTTTCCACTTCTGTAAAATGCACTGCATCTGAAACAAGTCTTCAAA
Tetraodon   1 ACCACTCTCACTGATCTCTCTTTTTCATGAACTGCGATGCATATGAAACAAGTCTTCAAA
Fugu        1 ACCACTCTCACTGATCTCTCTTTTTTCATGAACTGCCATGCATATGAAACAAGTCTTCAAA
Danio       1 AACGCTCTCAGTGATCTGCCTCTACCA----CGCCGAAC----TGAAACAAGTCTTCAAA
                                                         =============================
Human      61 AAA--CAAGGCCTAAAAAAA--TCTACT-ACTCAGGGTGCCTTTGAGGATATTTAATTAA
Mouse      60 AAA--CACGGCCTAAAAAAAAATCTACT-ACTCAGGGTGCCTTTGAGGATATTTAATTAA
Rat        61 AAA--CAGGGCCT-AAAAAA--TCTACT-ACTCAGGGTGCCTTTGAGGATATTTAATTAA
Chicken    61 AAAAACAAGGCCTAAAAAAA---CTACT-ACTCAGGGTGCCTTTGAGGATATTTAATTAA
Xenopus    61 AAA--CAAGGCCTAAAAAAAA--CTACT-ACTCAGGGTGCCTTTGAGGATATTTAATTAA
Tetraodon  61 A----CAAGGCTTTAAAAAC--T-TACCCAGTCAGGGTGCCTTTGAGGATATTTAATTAA
Fugu       61 A----CAAGGCTTTAAAAAC--T-TACCCAGTCAGGGTGCCTTTGAGGATATTTAATTAA
Danio      53 A----CAAGGCTTTAAAAAC--T-TACT-AGTCAGGGTGCCTTTGAGGATATTTAATTAA
                                      V Insert1
              ===========================================================
Human     116 AATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATG
Mouse     117 AATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATG
Rat       115 AATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATG
Chicken   117 AATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATG
Xenopus   116 AATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATG
Tetraodon 114 AATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATG
Fugu      114 AATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATG
Danio     105 AATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATG
              ========================================================
Human     176 GATTCTCATTTGCATATTGCATACAATTCATCAATTACTCAAGTATGAAAGGAGCACATT
Mouse     177 GATTCTCATTTGCATATTGCATACAATTCATCAATTACTCAAGTATGAAAGGAGCACATT
Rat       175 GATTCTCATTTGCATATTGCATACAATTCATCAATTACTCAAGTATGAAAGGAGCACATT
Chicken   177 GATTCTCATTTGCATATTGCATACAATTCATCAATTACTCAAGTATGAAAGGAGCACATT
Xenopus   176 GATTCTAATTTGCATATTGCATACAATTCATCAATTACTCAAGTATGAAAGGAGCACATT
Tetraodon 174 GATTCTCATTTGCATATTGCATACAATTCATCAATTACTCAAGTATGAAAGGAGCGCATT
Fugu      174 GATTCTCATTTGCATATTGCATACAATTCATCAATTACTCAAGTATGAAAGGAGCGCATT
Danio     165 GATTCTCATTTGCATATTGCATACAATTCATCAATTACTCAAGTATGAAAGGAGCACATT
                   V Insert2
Human     236 TCCCTTGGAGCTGCCTGCTACCCTGCCAACATTTGAAATGAGGGAAAGAGCAAGACTGTC
Mouse     237 TCCCTTGGAGCTGCCTGCTACCCTGCCAACATCTGAAATGAGGGAAAGAGCAAGACTGTC
Rat       235 TCCCTTGGAGCTGCCTGCTACCCTGCCAACATCTGAAATGAGGGAAAGAGCAAGACTGTC
Chicken   237 TCCCTTGGAGCTGCATGCTACCCTGCCAACATTTGAAATGAGGGAAAGAGCAAGACTGTC
Xenopus   236 TCCCTTGGAGCTGCCTGCTACTCTGCCAACATTTGAAATGAGGAAAAGAGCAAGACTGTC
Tetraodon 234 TCCCTTGGAGCTGCCTGCTAGCCTGCCAACATTTGAAATGAGAGAAAGAGCACGACTGTC
Fugu      234 TCCCTTGGAGCTGCCTGCTAGCCTGCCAACATTTGAAATGAGAGAAAGAGCACGACTGTC
Danio     225 TCCCTTGGAGCTGCCTGCTGCCCTGCCAGCATTTGAAATGAGAGAAAGAGCGATACTGTC

Human     296 AGGCATTCACA-CAAACTTTCTTCCA-AATGTCTGCTCCTTGATTAATCTAATTTTCTAG
Mouse     297 AGGCATTCACA-CAAACTTTCTTCCA-AATGTCTGCTCCTTGATTAATCTAATTTTCTAG
Rat       295 AGGCATTCACA-CAAACTTTCTTCCA-AATGTCTGCTCCTTGATTAATCTAATTTTCTAG
Chicken   297 AGGCATTCACA-CAAACTTTCTTCCA-AATGCCTGCTCCTTGATTAATCTAATTTTCTAG
Xenopus   296 AGCCGTTCGCA-CAAACTTTCTTCTT-AATGTCTGCTCCTTGATTAATCTAATTTTCTAG
Tetraodon 294 AGGCATTCACCGCAAACTTTTCCCCACAATGTCTGTGCGCTGATTAATCTCATTTTCTAG
Fugu      294 AGGCATTCACCGCAAACTTTTCCCCACAATGTCTGTGCGCTGATTAATCTCATTTTCTAG
Danio     285 AGGCATTCCAAACAAACTTT-CCCCACAATGTCTGCCTCTTGATTAATCTAATTTCCAAG

Human     354 ATCTTCCCT-ACAAGATACACCAACAGCCCTGGTGCACATTTCTATTATTTCGCCTGTCT
Mouse     355 ATATTCCCT-ACAAGATATAACAACAGCTCTAA-GCACATTTCTATTATTTCTCCTGTCT
Rat       353 ATATTCCCT-ACAAGATATACTAACAGCTCTG-TGCACATTTCTATCATTTCTTCTGTCC
Chicken   355 ATATTCCTT-ACAAGATACAACAACAGCTCTTCTGCAAATTTCTATTATTTCTCCTGTCT
Xenopus   354 CCATTTCTT-ACAAGATCCAACAACAGCCTTTCTGCAACTTTCAATTATTCTTGCTGTCT
Tetraodon 354 GCATCCATT-ACAAGATGTACA-CAGTCCAGCTGCACTTTTCAATTATTTCTCCTGTCC
Fugu      354 GCATCCCTT-ACAAGATGTACA-CAGTCCAGCTGCACTTTTCAATTATTTCTCCTGTCC
Danio     344 GCATCCTTTTACAAGATTGTACA-CAGTCCAGCTGCAGTTTTCAATTATATGTCCTGTCC

Human     413 TCATCACCAAAA
Mouse     413 TCATCACCAAAA
Rat       411 TCACCCCCCCCC
Chicken   414 TCATCATAAAAT
Xenopus   413 TTATTGCAAAAT
Tetraodon 412 TCTCAACTAAGT
Fugu      412 TCTCAACTAAGT
Danio     403 TCACCACTAACT
```
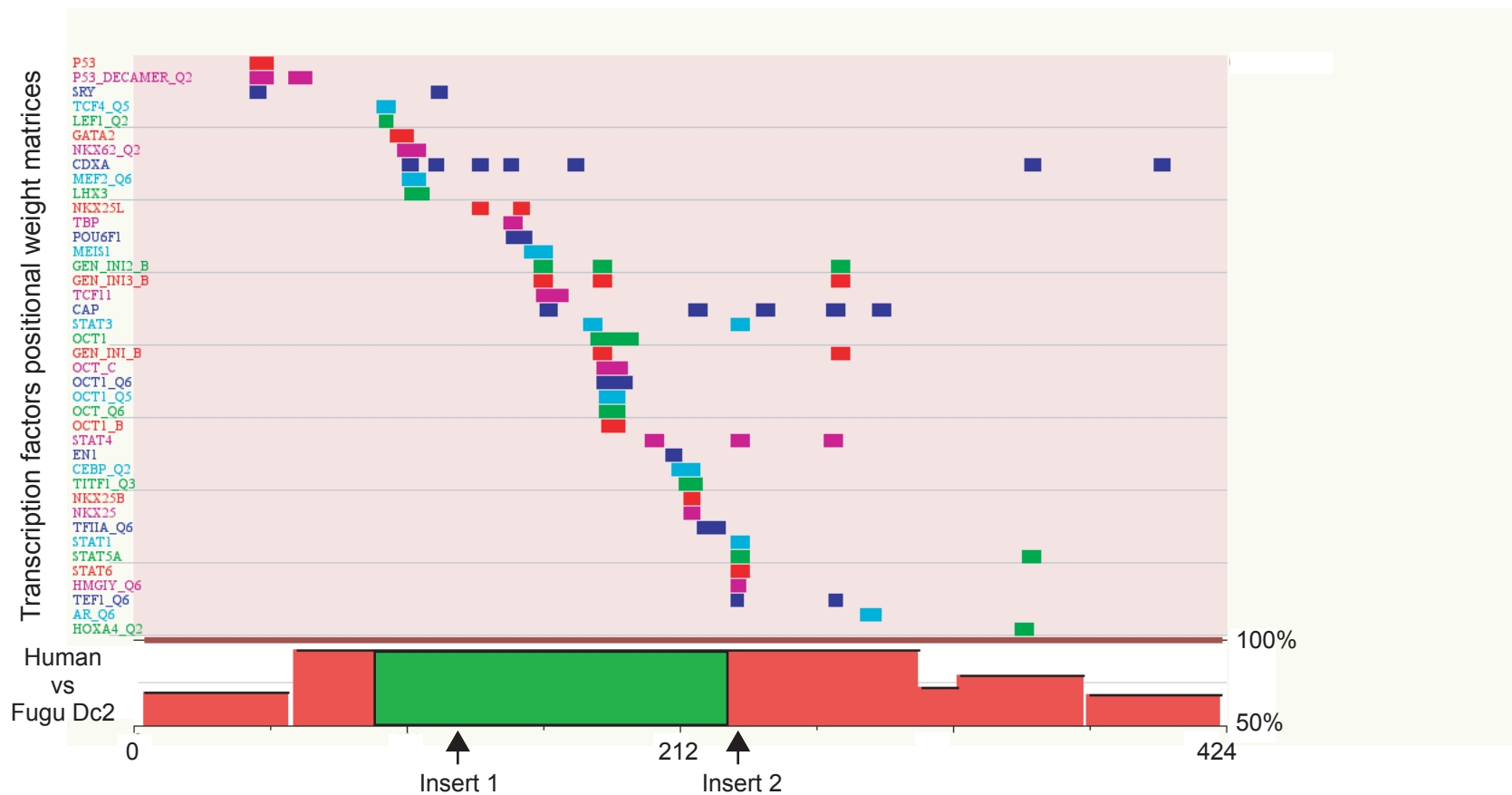
Figure 3

Figure 3 - Poulin et al.

Figure 4

Figure 4

```
87   TCAGGGTGCCTTTGAGGATATTTAATTAAAATCTAATCCTGCATTCAT  Human DACH Dc2
1    TCAGAGCGCCTTTGAAGATACTTAATTAAAATCTAATACCACATTCTT  Human DACH2 Dc2'


135  TAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATGGATTCTC  Human DACH Dc2
49   AAAAGCTCACATAAATTAAGCTGTCATTCAGGAGATTTATGCATTCAC  Human DACH2 Dc2'

183  ATTTGCATATTGCATACAATTCATCAATTACTCAAGTATGAAAGGAGC  Human DACH Dc2
97   ATTTGCATATTACATACAATTCATCAATTACTCATGTTTGAAAGCAAT  Human DACH2 Dc2'
```
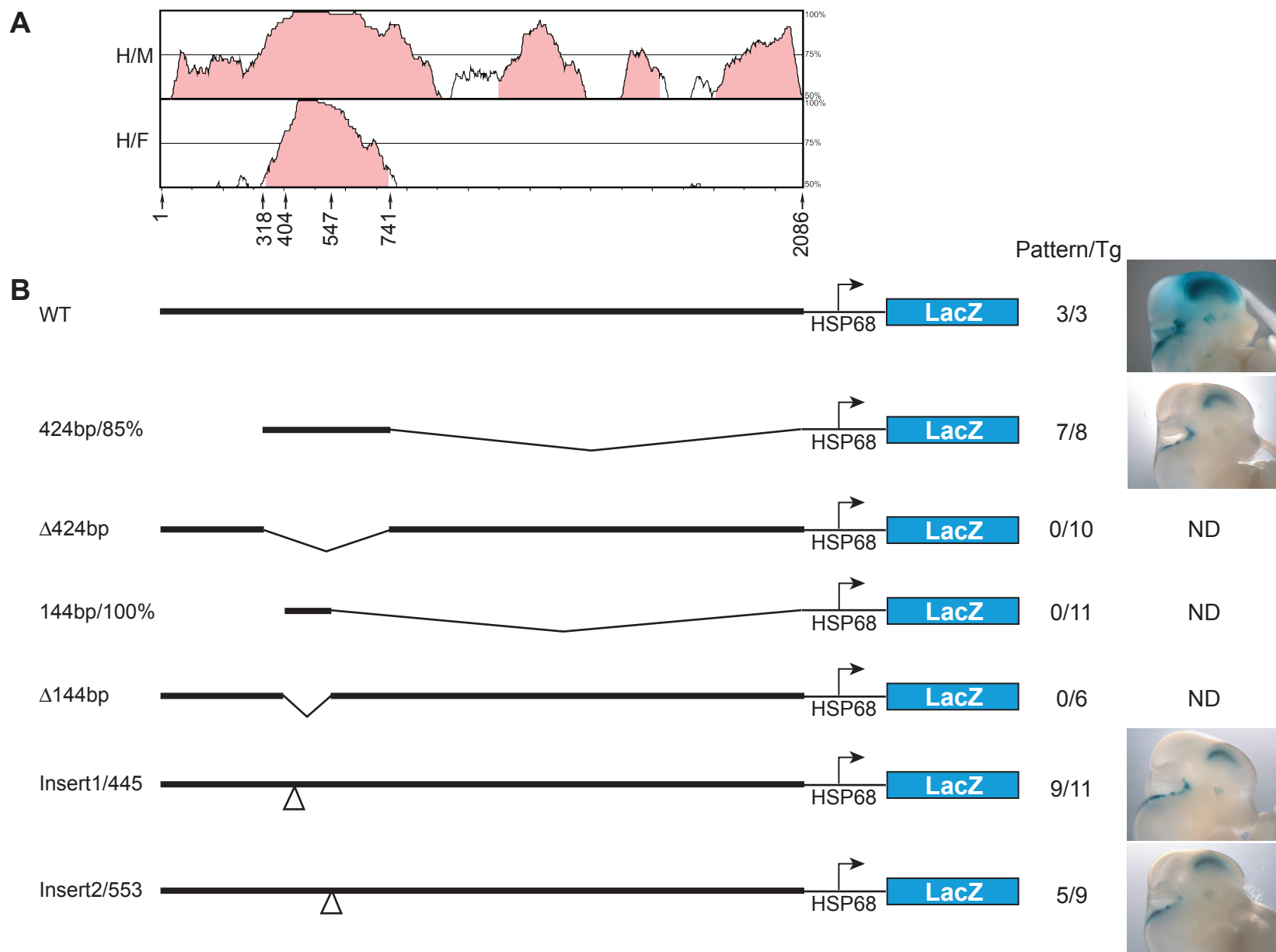
Figure 5

Figure 5 - Poulin et al.