



**Conservation of RET Regulatory Function from
Human to Zebrafish Without Sequence Similarity**

Shannon Fisher, *et al.*
Science **312**, 276 (2006);
DOI: 10.1126/science.1124070

**The following resources related to this article are available online at
www.sciencemag.org (this information is current as of June 14, 2007):**

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/312/5771/276>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/1124070/DC1>

This article **cites 26 articles**, 12 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/312/5771/276#otherarticles>

This article has been **cited by** 10 article(s) on the ISI Web of Science.

This article has been **cited by** 7 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/312/5771/276#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

of the fatty acid synthetase complex that act in trans.

Very recently, the crystal structures of the mammalian and fungal fatty acid synthetases were published that contain cis-acting ACP domains (25, 26). In both crystal structures, the ACP domains are not visible, most likely due to their inherent dynamics. These structures, however, also show that, in addition to the conformational flexibility (23) the entire ACP domain has to move to bridge the distance between the individual active sites. In contrast to the fatty acid synthase (FAS) system, where one ACP domain has to shuttle the substrates between multiple and distantly located catalytic centers, the PCP domains of NRPS systems transfer the substrate between an A-domain and the two neighboring C-domains, thus potentially making the intradomain dynamics sufficient for substrate shuttling. For a detailed understanding of the molecular mechanism of the FAS, polyketide synthase, and NRPS systems, the characterization of the intradomain and interdomain dynamics of the PCP and ACP domains will be crucial.

References and Notes

- C. T. Walsh, *Science* **303**, 1805 (2004).
- S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **105**, 715 (2005).
- T. Stachelhaus, A. Hüser, M. A. Marahiel, *Chem. Biol.* **3**, 913 (1996).
- H. D. Mootz, D. Schwarzer, M. A. Marahiel, *ChemBioChem* **3**, 490 (2002).
- T. Stein, *J. Biol. Chem.* **271**, 15428 (1996).
- T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure* **8**, 407 (2000).
- W. F. van Gunsteren, P. H. Hünenberger, A. E. Mark, P. E. Smith, I. G. Tironi, *Comput. Phys. Commun.* **91**, 305 (1995).
- H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comput. Phys. Commun.* **91**, 43 (1995).
- E. R. P. Zuiderweg, *Biochemistry* **41**, 1 (2002).
- R. H. Lambalot *et al.*, *Chem. Biol.* **3**, 923 (1996).
- K. Reuter, M. R. Mofid, M. A. Marahiel, R. Ficner, *EMBO J.* **18**, 6823 (1999).
- M. R. Mofid, R. Finking, M. A. Marahiel, *J. Biol. Chem.* **277**, 17023 (2002).
- D. Schwarzer, H. D. Mootz, U. Linne, M. A. Marahiel, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14083–14088 (2002).
- R. Finking, M. R. Mofid, M. A. Marahiel, *Biochemistry* **43**, 8946 (2004).
- M. R. Mofid, R. Finking, L. O. Essen, M. A. Marahiel, *Biochemistry* **43**, 4128 (2004).
- A. T. Brünger *et al.*, *Acta Crystallogr.* **54**, 905 (1998).
- C. Dominguez, R. Boelens, A. M. J. J. Bonvin, *J. Am. Chem. Soc.* **125**, 1731 (2003).
- T. Mittag, B. Schaffhausen, U. L. Günther, *Biochemistry* **42**, 11128 (2003).
- C. S. Goh, D. Milburn, M. Gerstein, *Curr. Opin. Struct. Biol.* **14**, 104 (2004).
- D. Kern, E. R. P. Zuiderweg, *Curr. Opin. Struct. Biol.* **13**, 748 (2003).
- H. C. Wong, G. Liu, Y. M. Zhang, C. O. Rock, J. Zheng, *J. Biol. Chem.* **277**, 15874 (2002).
- M. A. C. Reed *et al.*, *Org. Biomol. Chem.* **1**, 463 (2003).
- Y. Kim, J. H. Prestegard, *Biochemistry* **28**, 8792 (1989).
- M. Andreu, R. B. Hill, J. H. Prestegard, *Protein Sci.* **4**, 983 (1995).
- T. Maier, S. Jenni, N. Ban, *Science* **311**, 1258 (2006).
- S. Jenni, M. Leibundgut, T. Maier, N. Ban, *Science* **311**, 1263 (2006).
- G. M. Morris, D. S. Goodsell, R. Huey, A. J. Olson, *J. Comput. Aided Mol. Des.* **10**, 293 (1996).
- This work was supported by the Centre for Biomolecular Magnetic Resonance Frankfurt (BMRZ) and DFG grants BE 19-11 and MA 811/19-1. Coordinates of the A state (2GDY), the A/H state (2GDW), the H state (2GDY), and the complex of the cofactor CoA with Sfp (2GEO) and of the complex of Sfp with CoA, Mg²⁺ and the A state of the apo-PCP (2GE1) have been deposited in the Protein Data Bank.

Supporting Online Material

www.sciencemag.org/cgi/content/full/312/5771/273/DC1
Materials and Methods
Figs. S1 to S6
Table S1

22 November 2005; accepted 20 March 2006
10.1126/science.1122928

Conservation of *RET* Regulatory Function from Human to Zebrafish Without Sequence Similarity

Shannon Fisher,^{1,2*} Elizabeth A. Grice,^{1*} Ryan M. Vinton,¹ Seneca L. Bessling,¹ Andrew S. McCallion^{1,3†}

Evolutionary sequence conservation is an accepted criterion to identify noncoding regulatory sequences. We have used a transposon-based transgenic assay in zebrafish to evaluate noncoding sequences at the zebrafish *ret* locus, conserved among teleosts, and at the human *RET* locus, conserved among mammals. Most teleost sequences directed *ret*-specific reporter gene expression, with many displaying overlapping regulatory control. The majority of human *RET* noncoding sequences also directed *ret*-specific expression in zebrafish. Thus, vast amounts of functional sequence information may exist that would not be detected by sequence similarity approaches.

A current hypothesis is that sequences conserved over greater evolutionary distances are more likely to be functional than those conserved over lesser distances (1). Many recent publications have focused attention on the regulatory potential of “ultra-conserved” noncoding sequences, conserved across great evolutionary distances, e.g., human to fugu (2–9) [≥ 300 million years, or average 74% protein identity (10)]. These are frequently enhancers

associated with developmental genes, consistent with strong selective pressure to preserve critical mechanisms. Analyses of identified sequences have generally fallen into two categories: analyses confined to mammals, with functional verification done in mice, or analyses including mammalian and teleost sequences, focusing on highly conserved sequences alignable at the extremes. However, simply because an expression pattern is preserved through evolution, it does not necessarily follow that the cis-regulatory elements controlling that expression in one species will function in a second.

We have explicitly tested two hypotheses: First, using selective pressure as a guide across moderate evolutionary distances, we can identify the majority of enhancers control-

ling expression at a particular locus by functional testing in a comprehensive, unbiased manner, and second, regulatory function of noncoding sequences will be conserved over evolutionary distances beyond the limit of overt sequence conservation.

We have focused on the regulatory control of the gene encoding the RET receptor tyrosine kinase. *RET* is expressed in neural crest, urogenital precursors, adrenal medulla, and thyroid during embryogenesis, and in specific central and peripheral neurons and endocrine cells during development and postnatally (11). Although *RET* expression is highly conserved across evolution (12–15), only the exons encoding the tyrosine kinase domain are overtly conserved [$\geq 70\%$, ≥ 100 base pairs (bp)] from humans to zebrafish (16–18). We first compared the genomic sequence of a ~ 200 -kilobase (kb) segment encompassing the zebrafish *ret* gene with the orthologous interval in fugu (Fig. 1), using AVID/VISTA (19, 20). We generated 10 ZCS (zebrafish conserved sequence) amplicons, corresponding to 14 discrete noncoding sequences (table S1).

We also used these criteria to identify conserved noncoding human sequences, comparing a ~ 200 -kb segment encompassing human *RET* with the orthologous genomic intervals in 12 nonhuman vertebrates (16). We selected sequences shared among human and at least three nonprimate mammals (21). In total 13 HCS (human conserved sequence) amplicons, encompassing 28 discrete conserved sequences (table S2) were generated for analysis.

¹McKusick–Nathans Institute of Genetic Medicine, ²Department of Cell Biology, ³Department of Comparative Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: sfisher@jhmi.edu (S.F.); amccalli@jhmi.edu (A.S.M.)

Although zebrafish transgenesis has been used to evaluate the regulatory potential of conserved noncoding sequences (2, 7, 22), its efficacy is compromised by mosaicism in injected (G_0) embryos. We developed a reporter vector based on the Tol2 transposon; reporter expression in G_0 embryos, driven from the ubiquitous *efla* promoter, was extensive and was dependent on transposase RNA (23).

All but one ZCS amplicon drove reporter expression consistent with endogenous *ret* expression (Table 1). As in the mouse, zebrafish *ret* is expressed in sensory neurons of the cranial ganglia, motor neurons in the ventral hindbrain, cells of the hypothalamus and pituitary primordia, sensory and motor neurons in the spinal cord, and primary sensory neurons in the olfactory pit (13, 14). We discovered elements driving expression consistent with all of these cell populations (Table 1), including small groups of cells, e.g., olfactory neurons (Fig. 2A) and lateral line placode ganglion (Fig. 3, A and B). Although *ret* is also expressed in amacrine and horizontal cell layers of the retina, we did not detect expression in the retina of G_0 embryos with any of the tested elements.

We found significant redundancy in the control of *ret* expression in the pronephric duct (Table 1; Fig. 2, C and D). Five elements drove expression in the intermediate mesoderm or pronephric duct; one was responsible for transient early expression (Fig. 2C), one for expression in the distal duct after 3 days (Fig. 2D), and three apparently redundantly control expression in the intervening period. Although three amplicons lie within a 5-kb region upstream of *ret*, they function independently in our assay. Similarly all but two ZCS amplicons drove expression in one or more cell populations of the central nervous system (Table 1), wherein *ret* is also dynamically expressed.

Surprisingly, 11 out of 13 HCS amplicons drove expression in cell populations consistent with zebrafish *ret* (Table 1). These included cells not present in mammals, such as the afferent neurons of the lateral line ganglia. We also observed multiple sequences driving expression in the excretory system, despite its developmental and anatomical differences between fish and mammals (Fig. 2G). Two sequences contained within a genomic interval deleted from the rodent lineage also functioned in zebrafish, in one case driving expression in the pituitary (Figs. 2E, 3E). Several pairs of elements drove similar expression patterns, despite lack of detectable sequence conservation (Table 1). To rule out the possibility that nonconserved sequences could fortuitously display enhancer activity, we analyzed expression from vectors containing nonconserved zebrafish ($n = 5$) or human ($n = 3$)

genomic DNA, from the *RET* intervals (tables S1 and S2). None of these nonconserved sequences provided reproducible patterns of expression.

Through analysis of G_0 expression, we identified enhancers active in small cell populations such as the cranial ganglia and olfactory neurons (Fig. 2), suggesting that

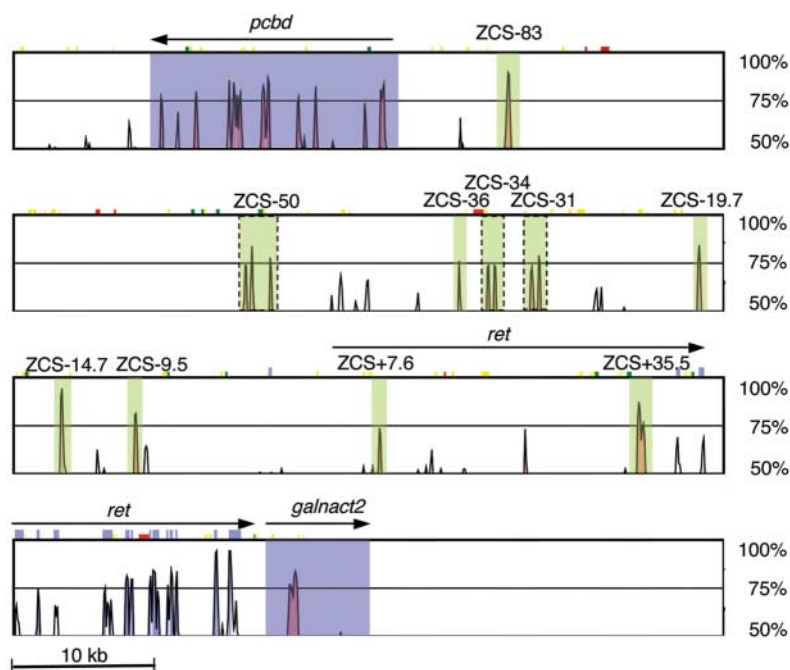


Fig. 1. Comparative sequence analysis of teleost *ret* loci reveals putatively functional noncoding sequences. VISTA plot displaying the alignment of the zebrafish *ret* locus with the orthologous fugu region. Red peaks represent conserved noncoding sequences; shaded green boxes represent ZCS amplicons. Boxes bordered by dashed lines denote amplicons containing two or more conserved sequences. *ret* exons are denoted by blue peaks. Red peaks boxed and shaded in blue denote 5' and 3' flanking genes *pcbd* and *galnact2*, respectively.

Table 1. Noncoding sequences from zebrafish *ret* or human *RET* direct expression consistent with endogenous *ret*. The elements are described by their species of origin and distance in kilobases from the translation start site, and (i.e., ZCS-50, HCS+16). Abbreviations: CG, cranial ganglia; SC, spinal cord; PND, pronephric duct; IM, intermediate mesoderm; NTC, notochord; OLF, olfactory pit/placode; +, present.

Constructs	Brain	SC	CG	ENS	NTC	OLF	Retina	Heart	IM/PND	Fin bud
ZCS-83	+	+	+			+				
ZCS-50	+	+	+		+			+		+
ZCS-36	+								+	
ZCS-34	+								+	
ZCS-31	+								+	
ZCS-19.7	+	+	+	+			+		+	
ZCS-14.7	+	+								
ZCS-9.5	+	+	+							
ZCS+7.6									+	
ZCS+35.5		+	+	+				+		
HCS-32	+	+	+					+		+
HCS-30									+	
HCS-23					+					
HCS-12		+							+	
HCS-8.7									+	
HCS-7.4										
HCS-5.2	+					+			+	
HCS+9.7				+					+	
HCS+16	+									
HCS+19	+	+								

*Expression before 24 hours.

mosaicism is not a significant limitation. We have passed a subset of transgenes through the germline (Fig. 3, A to C and E to G), to directly compare expression in G_0 and G_1 embryos. Expression of each transgene was largely consistent with that observed in G_0 phases (Fig. 3, A and B), although in some cases we observed additional expression, particularly in small groups of cells and at later time points [retina (Fig. 3G)]. We also evaluated many G_1 embryos using in situ hybridization (ISH) to detect *gfp* transcripts, which confirmed that green fluorescent protein (GFP) signal was present in *ret* positive cells (Fig. 3, C and D).

While still functioning as tissue-specific enhancers in zebrafish, some HCSs directed expression differing in timing or location from that of the endogenous *ret* gene. For example, HCS-32 drives GFP expression in dorsal spinal cord neurons, apparent between embryonic day 2 and 3. ISH analyses of G_1 transgenic embryos revealed expression at earlier stages in the posterior neural plate, where *ret* is not normally expressed. Additionally, two elements, HCS-23 and ZCS-50,

directed expression strongly to the notochord, again not a site of endogenous *ret* expression. One possible reason for these discrepancies is that we are assaying elements out of context. Also, physical proximity does not mean that these elements normally regulate *ret* expression. In the case of HCSs, individual transcription factor-binding sites (TFBSs) may have evolved sufficiently to display different functions (i.e., binding related proteins, binding with different affinity), reflected in altered regulatory activity of the element as a whole.

HCS function in zebrafish may arise from sequence elements ≤ 100 bp that are conserved but fail to meet our original criteria for identification. Consequently, we repeated our sequence analysis with AVID/VISTA, reducing the window size to 30 bp. We also analyzed the *RET* orthologous intervals using the anchored alignment algorithms Multi-LAGAN and Shuffle-LAGAN (24), the latter designed to detect alignable sequences in the presence of inversions and rearrangements. We also attempted to align each *RET* HCS independently, in both orientations, with the zebrafish *ret* interval (25). All analyses failed

to detect sequences alignable between human and zebrafish *RET* intervals. We further searched the entire zebrafish genome (26) for homologies to the examined HCSs. Sixty-five sequences within these HCSs of ≥ 20 nucleotides in length demonstrated $\geq 70\%$ identity with nonorthologous, intergenic zebrafish sequences, within 100 kb of a known or predicted gene; 41 out of 65 contain conserved TFBS motifs (table S3). However, we also aligned the nonconserved HCSs with the zebrafish genome and found alignments containing TFBSs at a similar frequency, which suggested that such analyses are not predictive of regulatory function. We posit that the responsible functional components in the conserved elements are single or multiple TFBSs (4 to 20 bp), beyond the ability of our current in silico tools to reliably detect. Our data suggest that restricting in vivo functional analyses to sequences conserved over great evolutionary distances (e.g., human to teleost) detects only a small fraction of functional information in the genome.

We have developed an efficient method to evaluate putative enhancer elements, allowing rapid assessment of in vivo function in a vertebrate embryo. This method is suitable for rapid screening of putative enhancers on a large scale, even where the orthologous zebrafish sequence is not available. Our approach represents a significant advance over previous methods because of the decreased mosaicism and improved germline transmission achieved with Tol2 vectors. The transparent external development of zebrafish facilitates dynamic analysis of reporter activity throughout embryogenesis, allowing detection of biological activity throughout development. This has allowed us to survey without bias all conserved sequences at a single, complex locus.

Our data strongly suggest that functional information is conserved in vertebrate sequences at levels below the radar of large-scale genomic sequence alignment, consistent with prior anecdotal observations (27, 28). Two alternative models could be invoked to

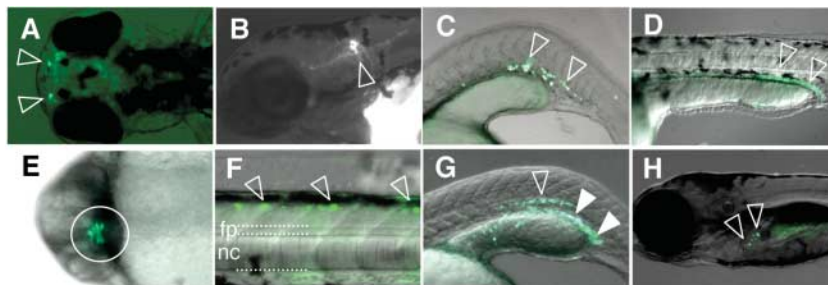
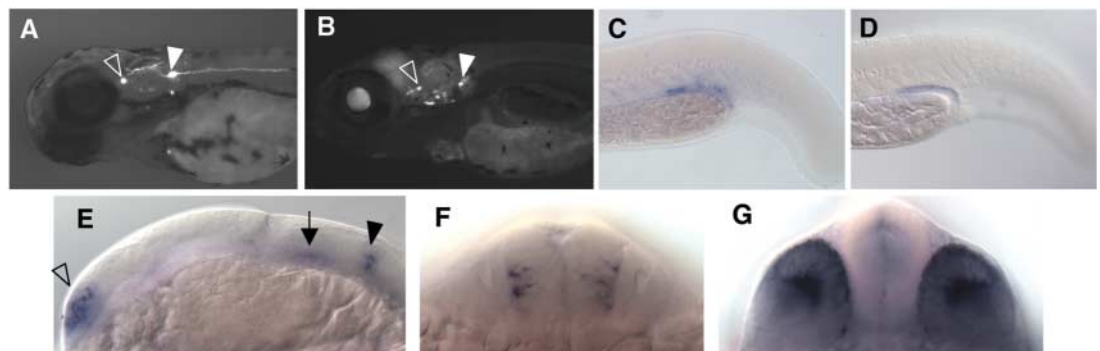


Fig. 2. Conserved noncoding sequences at the zebrafish and human *ret* loci drive reporter expression in zebrafish embryos consistent with the endogenous gene. Shown are GFP expression patterns in representative G_0 embryos. (A to D) Zebrafish elements drive expression in (A) bilateral olfactory pits (arrowheads; ZCS-83); (B) hindbrain neuron consistent with nVII facial motor neuron (arrowhead; ZCS-19.7); (C) pronephric duct before 24 hours (arrowhead; ZCS-34); and (D) pronephric duct at 3 days (arrowheads; ZCS-7.6). Human elements drive expression in (E), pituitary (encircled, HCS+16); (F) dorsal spinal cord neurons (arrowheads, HCS-32; fp, floor plate; nc, notochord); (G) pronephric duct (solid white arrowheads) and enteric neurons (open arrowhead; HCS+9.7); and (H) enteric neurons (open arrowheads, HCS+9.7).

Fig. 3. Mosaic G_0 expression accurately reflects expression in G_1 fish. (A) ZCS-35.5 G_0 embryos display GFP in cells of the anterior (open arrowhead) and posterior (solid white arrowhead) lateral line placode ganglia. (B) ZCS-35.5 G_1 embryos display GFP in the anterior (open arrowhead) and posterior (solid white arrowhead) lateral line placode ganglia, as in (A). (C) GFP detected by in situ hybridization (ISH) in the distal pronephric duct of ZCS+7.6 G_1 embryo at 24 hours, consistent with *ret* expression at the same stage (D). (E and F) GFP detected by ISH in the pituitary (open arrowhead), trigeminal



nuclei (arrow), and migrating nVII facial motor neurons [arrowhead in (E) and (F)] of a HCS+16 G_1 embryo. (G) GFP detected by ISH in the retina of G_1 ZCS-19.7 embryo.

explain our data. First, overall similar expression of the *RET* genes could be achieved through assemblage of analogously acting, although not orthologous, enhancers. A second, more parsimonious, explanation is that orthologous enhancer elements control expression of both *RET* genes, but have evolved beyond recognition through small changes in TFBSs, rearrangement of sites within enhancers, or multiple coevolved changes. Examination of enhancer evolution in *Drosophila* species reveals examples of these types of sequence changes, confounding traditional sequence alignment approaches while preserving enhancer function across species (29–31). Comparison of human and mouse enhancer sequences suggests that similar widespread turnover of TFBSs is observed in vertebrate evolution (28), although there is no corresponding functional data to confirm that such changes occur while preserving the function of the enhancers. Our data cannot distinguish between these two models; however, it must be the case that largely the same set of transcription factors regulate expression of either gene, and the binding of these is conserved from mammalian to teleost enhancer elements, which allows the HCSs to function in zebrafish. These data may now significantly alter the manner in

which the biological relevance of vertebrate noncoding sequences is evaluated.

References and Notes

1. D. Boffelli, M. A. Nobrega, E. M. Rubin, *Nat. Rev. Genet.* **5**, 456 (2004).
2. A. Woolfe *et al.*, *PLoS Biol.* **3**, e7 (2005).
3. M. A. Nobrega, I. Ovcharenko, V. Afzal, E. M. Rubin, *Nature* **302**, 413 (2003).
4. S. Bagheri-Fam, C. Ferraz, J. Demaille, G. Scherer, D. Pfeifer, *Genomics* **78**, 73 (2001).
5. N. Baroukh *et al.*, *Mamm. Genome* **16**, 91 (2005).
6. F. Poulin *et al.*, *Genomics* **85**, 774 (2005).
7. E. de la Calle-Mustienes *et al.*, *Genome Res.* **15**, 1061 (2005).
8. A. Sandelin *et al.*, *BMC Genomics* **5**, 99 (2004).
9. G. Bejerano *et al.*, *Science* **304**, 1321 (2004).
10. V. Veeramachaneni, W. Makalowski, *Nucleic Acids Res.* **33**, D442 (2005).
11. A. S. McCallion, A. Chakravarti, in *Inborn Errors of Development*, C. Epstein, R. Erikson, A. Wynshaw-Boris, Eds. (Oxford Univ. Press, Oxford, 2004).
12. M. Hahn, J. Bishop, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1053 (2001).
13. C. V. Marcos-Gutierrez, S. W. Wilson, N. Holder, V. Pachnis, *Oncogene* **14**, 879 (1997).
14. B. W. Bigrove, D. W. Raible, V. Walter, J. S. Eisen, D. J. Grunwald, *J. Neurobiol.* **33**, 749 (1997).
15. V. Pachnis, B. Mankoo, F. Costantini, *Development* **119**, 1005 (1993).
16. E. S. Emison *et al.*, *Nature* **434**, 857 (2005).
17. A. S. McCallion *et al.*, *Cold Spring Harb. Symp. Quant. Biol.* **68**, 373 (2003).
18. C. S. Kashuk *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8949 (2005).
19. K. A. Frazer, L. Pachter, A. Poliakov, E. M. Rubin, I. Dubchak, *Nucleic Acids Res.* **32**, W273 (2004).
20. Materials and methods are available as supporting material on Science Online.
21. E. A. Grice, E. S. Rochelle, E. D. Green, A. Chakravarti, A. S. McCallion, *Hum. Mol. Genet.* **14**, 3837 (2005).
22. S. Ellingsen *et al.*, *Development* **132**, 3799 (2005).
23. S. Fisher, E. A. Grice, R. M. Vinton, S. L. Bessling, A. S. MacCallion, unpublished observations.
24. See http://lagan.stanford.edu/lagan_web/index.shtml.
25. See <http://genome.ucsc.edu/cgi-bin/hgBlat>.
26. DanRer2 (www.sanger.ac.uk/Projects/D_eriol/), accessed June 2004.
27. B. Gottgens *et al.*, *Nat. Biotechnol.* **18**, 181 (2000).
28. L. A. Pennacchio *et al.*, *Science* **294**, 169 (2001).
29. B. P. Berman *et al.*, *Genome Biol.* **5**, R61 (2004).
30. M. Z. Ludwig, C. Bergman, N. H. Patel, M. Kreitman, *Nature* **403**, 564 (2000).
31. M. Z. Ludwig *et al.*, *PLoS Biol.* **3**, e93 (2005).
32. E. T. Dermitzakis, A. G. Clark, *Mol. Biol. Evol.* **19**, 1114 (2002).
33. The authors thank K. Kawakami for his generous gift of the pTZIG vector. This study was supported by a Basil O'Connor award from the March of Dimes (A.S.M.) and grant AR048101 from the NIH (S.F.). Zebrafish sequence data (DanRer2) were produced by the *Danio rerio* Sequencing Group at the Sanger Institute and can be obtained by ftp transfer (<ftp://ftp.ensembl.org/pub/assembly/zebrafish/Zv4release>).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1124070/DC1

Materials and Methods

Tables S1 to S3

References

19 December 2005; accepted 8 March 2006

Published online 23 March 2006;

10.1126/science.1124070

Include this information when citing this paper.

A Common Genetic Variant Is Associated with Adult and Childhood Obesity

Alan Herbert,^{1*} Norman P. Gerry,¹ Matthew B. McQueen,² Iris M. Heid,^{3,4} Arne Pfeufer,^{5,6} Thomas Illig,^{3,4} H.-Erich Wichmann,^{3,4,7} Thomas Meitinger,^{5,6} David Hunter,^{2,8,9} Frank B. Hu,^{2,8,9} Graham Colditz,^{8,9} Anke Hinney,¹⁰ Johannes Hebebrand,¹⁰ Kerstin Koberwitz,^{6,10} Xiaofeng Zhu,¹¹ Richard Cooper,¹¹ Kristin Ardlie,¹² Helen Lyon,^{13,14,15} Joel N. Hirschhorn,^{13,14,15} Nan M. Laird,¹⁶ Marc E. Lenburg,¹ Christoph Lange,^{9,13} Michael F. Christman^{1*}

Obesity is a heritable trait and a risk factor for many common diseases such as type 2 diabetes, heart disease, and hypertension. We used a dense whole-genome scan of DNA samples from the Framingham Heart Study participants to identify a common genetic variant near the *INSIG2* gene associated with obesity. We have replicated the finding in four separate samples composed of individuals of Western European ancestry, African Americans, and children. The obesity-predisposing genotype is present in 10% of individuals. Our study suggests that common genetic polymorphisms are important determinants of obesity.

Obesity is associated with an increased risk of type 2 diabetes mellitus, heart disease, metabolic syndrome, hypertension, stroke, and some forms of cancer (1). It is commonly assessed by calculating an individual's body mass index (BMI) [weight/(height)² in kg/m²] as a surrogate measurement. Individuals with a BMI \geq 25 kg/m² are classified as overweight, and those with a BMI \geq 30 kg/m² are considered obese. Having a BMI

over 25 kg/m² increases the risk of death (2). Presently, 65% of Americans are overweight and 30% are obese (3). Genetic factors contribute significantly to the etiology of obesity (4, 5), with estimates of the heritability of BMI ranging from 30 to 70% (6–9).

To identify common genetic variants associated with elevated BMI, we have studied individuals from the National Heart, Lung, and Blood Institute (NHLBI)–Framingham Heart

Study (FHS) (10). The participants were enrolled from the community without being selected for a particular trait or disease and were followed over 24 years (table S1). In this population, heritability estimates for BMI range between 37 and 54% (11, 12).

Using families from this sample, we performed a genome-wide association analysis, using a testing strategy for quantitative traits in

¹Department of Genetics and Genomics, Boston University Medical School, E613, 715 Albany Street, Boston, MA 02118, USA. ²Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA. ³Institute of Epidemiology, ⁴KORA Group, ⁵Institute of Human Genetics, Forschungszentrum für Umwelt und Gesundheit, D-85764 Neuherberg, Germany. ⁶Institute of Human Genetics, Technical University Munich, D-81671 Munich, Germany. ⁷Institute for Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany. ⁸Nurses Health Study, ⁹Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA. ¹⁰Department of Child and Adolescent Psychiatry of the University of Duisburg-Essen, D-45147 Essen, Germany. ¹¹Department of Preventive Medicine and Epidemiology, Loyola University Medical Center, Maywood, IL 60153, USA. ¹²Genomics Collaborative, SeraCare Life Sciences, Cambridge, MA, USA. ¹³Program in Genomics and Divisions of Genetics and Endocrinology, Children's Hospital, Boston, MA 02115, USA. ¹⁴Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. ¹⁵Broad Institute of Harvard and MIT, Cambridge, MA 02139, USA. ¹⁶Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: aherbert@bu.edu (A.H.); mfc@bu.edu (M.F.C.)