Title
Characterizing Regulatory Elements That Could Lead to Obesity Susceptibility

Author
Kim, Mee Jean

Publication Date
2013

Peer reviewed|Thesis/dissertation

Characterizing Regulatory Elements That Can Lead to Obesity Susceptibility

by

Mee Jean Kim

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION
of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

ACKNOWLEDGEMENTS

This dissertation is not only a reflection of the work I have accomplished during my graduate studies here at UCSF but a reflection of the foundation I received as I started a career in scientific research, the continual professional and personal support of my advisors, peers, friends, community and family around me. Without any one of these people, I would not be able to stand here today, having completed this degree.

One of my first thanks is to Dr. Gary Murray for taking me on as a Post-baccalaureate IRTA fellow at the National Institutes of Health; this is where my transition from being a chemistry major to human genetics began. This position and experience prepared me for graduate school.

I would like to acknowledge the UCSF Pharmaceutical Sciences and Pharmacogenomics program and my classmates. I thank the program for the many different opportunities in which I was able to participate throughout my graduate career: funding, retreat organizing and admissions to name a few. Many thanks to my classmates with whom I endured first year classes, second year qualifying exams, third-to-fifth year research slumps and finally, this: our dissertations.

My deepest respect and gratitude go to my advisor, Dr. Nadav Ahituv. It took a lot of faith taking me in (mostly in part that he was a new PI and I a first year graduate student with minimal molecular biology/genetics training) and I am thankful for the opportunity to grow into a scientist alongside of him. I am grateful him for the many different successful research opportunities, projects and collaborations in which I was able to participate. I admire his passion for science, his drive to succeed and his ability to balance career and family – all of which I hope to strive towards in my own career. I also

thank him for teaching me the hydrodynamic tail vein injection. A technique that has allowed me to participate in many projects and collaborations and a technique from which I cannot escape and, at this rate, most likely will be the reason of my future employments.

I would also like to thank my thesis committee, Drs. Christian Vaisse and Kaveh Ashrafi. I had the best thesis committee faculty; our committee meetings were actually fun. Not only was I fortunate to be surrounded by two very smart, encouraging and supportive scientists, their feedback and advice were invaluable during the course of my study, as were my memories of playing soccer with them.

I have worked with some of the best scientists and many of them were in the Ahituv Lab. I am indebted to Loan Nguyen, who trained and taught me the foundations of molecular biology and human genetics. I would also like to express my appreciation to Dr. Ramon Birnbaum; it would not have been the same experience without his openness, advice, encouragements and honesty; many thanks for allowing me to join him on his eExons projects. Special shout out to Nir Oksenberg. You made my mornings in lab bearable with updates on the current status of the Niners, Giants and sports in general. It was also fun to collaborate with you on this thesis project. Thank you so much for the millions of sections you generated for the millions of *in situ*s and LacZ stainings I performed. It has been a unique experience to learn the ins and outs of building a new lab from the ground up and I thank all the lab members past and present for an unforgettable 5.5 years; thank you for making our lab a great and fun place to conduct research.

On the subject of collaborations and help, I would also like to acknowledge Gabriel McKinsey from the Rubenstein lab for his technical support for *in situ*

hybridizations and the generous gift of the mouse *Sim1* probe; likewise with Dr. Kerstin Seidel from the Klein lab and her support in optimizing this assay, which ended up being a very important aspect of this project. I would also like to acknowledge Dr. Thomas Hoffmann who performed the 1000 Genomes Project bioinformatic analysis on SCE2.

I would also like to thank Dr. Yien Ming Kuo. I cannot fully express my gratitude and gratefulness for everything you have done and meant to me during this experience. I am truly indebted to you on the technical, professional and personal level. I cannot thank you enough and I am lucky to call you friend.

With all the support I received throughout my education, it would have not been the same without the generous encouragements of my friends throughout the process. Thank you to those who have been with me from the very beginning of my scientific career – even before my move to UCSF and to those here in the City who I have met and been able to draw upon for the "occasional" cheer up session or to celebrate milestones. Thank you so much for being my friends.

Among these friends, I would not be here today without the generosity, kindness, and spiritual support of my Lifehouse Covenant Church community. I would not have survived graduate school without you. Thank you all for doing life with me and for helping me grow in the six years I have been at LCC. A very special thank you to the Miki Family, who have generously opened up their home and lives for me to be a part of. I have been truly blessed.

I would also like to express my sincere gratitude to the Dolich Family, who basically adopted into their family since I have been in San Francisco; thank you for being a second family and home to go to on the holidays; a very special thank you to

Lindsey Dolich Felt. It has been an immense joy having your friendship all the way from Haverford to graduate school and I am truly grateful for your steadfast support.

Lastly, I would like to thank my family. Sophia and Andrew, you are the reason I strive to do my best every day. Mom and Dad, thank you for your sacrifices, unconditional support, encouragements, love and prayers. You are the reason why I want to give back to others. I dedicate this thesis work and my doctoral degree to you.

ABSTRACT

Haploinsufficiency of the Single Minded homology 1 (*SIM1*) gene in humans and mice leads to severe obesity, suggesting that altered expression of *SIM1*, by way of regulatory elements such as enhancers, could predispose individuals to obesity. To identify enhancers that could regulate *SIM1*, we used comparative genomics coupled with zebrafish and mouse transgenic enhancer assays. Due to the dual role of *Sim1* in hypothalamic development and in adult energy homeostasis, the enhancer activity of these sequences was annotated from embryonic to adult age. Of the seventeen tested sequences, two (SCE2 and SCE8) were found to have midbrain enhancer activity in zebrafish. Both SCE2 and SCE8 also exhibited embryonic hypothalamus enhancer expression in mice, and time course analysis of SCE2 activity showed overlapping expression to *Sim1* from embryonic to adult age. Using a deletion series, we identified the critical region in SCE2 that is needed for hypothalamus enhancer activity. Sequencing this region in obese and lean cohorts revealed a higher prevalence of SNPs that were unique to obese individuals, with one variant reducing developmental enhancer activity in zebrafish. In summary, we have characterized two hypothalamus enhancers in the *SIM1* locus and identified a set of obesity-specific SNPs within one of them, which may predispose individuals to obesity.

TABLE OF CONTENTS

List of Tables

List of Figures

Introduction

Results

INTRODUCTION

THE OBESITY EPIDEMIC

Obesity is a medical condition in which there is an excess proportion of accumulated body fat, to the extent that may be detrimental to one's health. This is the result of an energy homeostasis imbalance between energy intake and energy expenditure. The clinical definition of obesity is measured by body mass index (BMI) (developed by Belgian polymath Adolphe Quetelet), which is calculated by one's weight (kg) divided by the square of their height (m^2). A BMI of 25 or higher is considered overweight while a BMI of 30 or higher is categorized as obese, with graded severity as BMI increases.

In the last decade, obesity has reached epidemic rates of prevalence, especially in the United States. According to the National Center for Health Statistics, it is estimated that over one third (35.7%) of the U.S. population, or 78 million people, is currently obese compared to 27.5% in 1999-2000 (1) (Figure 1), with another third of the population considered overweight and comparable overweight/obesity rates (combined 47%) globally (2, 3) (Figure 2). Obesity is a major public health concern due to its contribution to increasing risk for other diseases such as cancer, cardiovascular disease, type 2 diabetes, stroke, sleep apnea, asthma and osteoarthritis (4–6); most recently, obesity was classified a disease by the American Medical Association (AMA) at the annual 2013 AMA meeting.

The cause of obesity has been attributed to a combination of increased sedentary lifestyle leading to reduced energy expenditure, dietary excess and other environmental factors; equally important is the underlying genetic predisposition that lays the foundation on which environmental and exogenous factors manifest their propensity to

drive the obesity phenotype. This dissertation focuses on the study of the genetic contributions to human obesity.

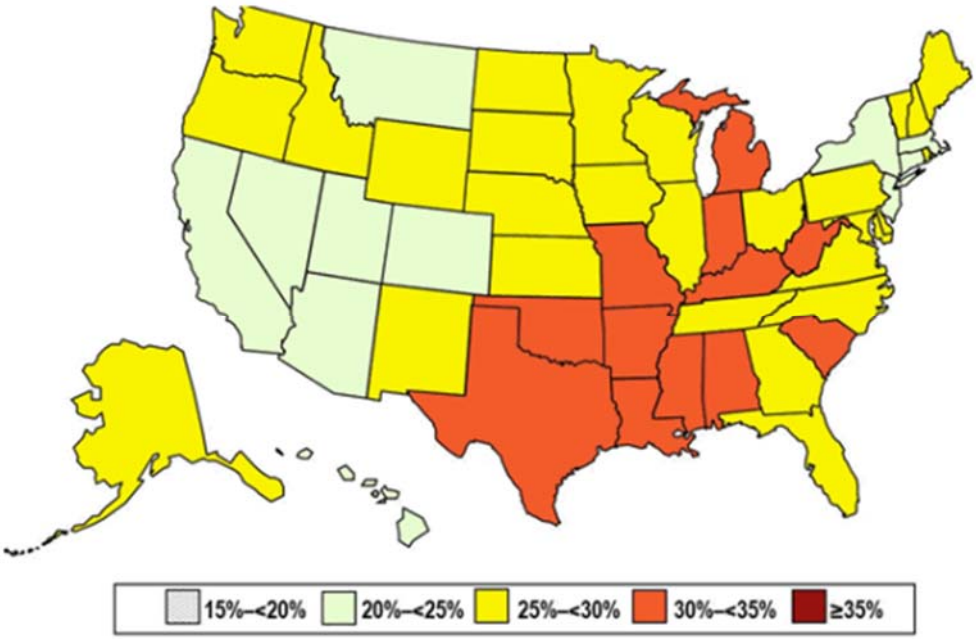

Figure 1. 2011 National prevalence of obesity. The prevalence of self-reported obesity in adults across the United States. Approximately every state has a prevalence of at least 20% (7).

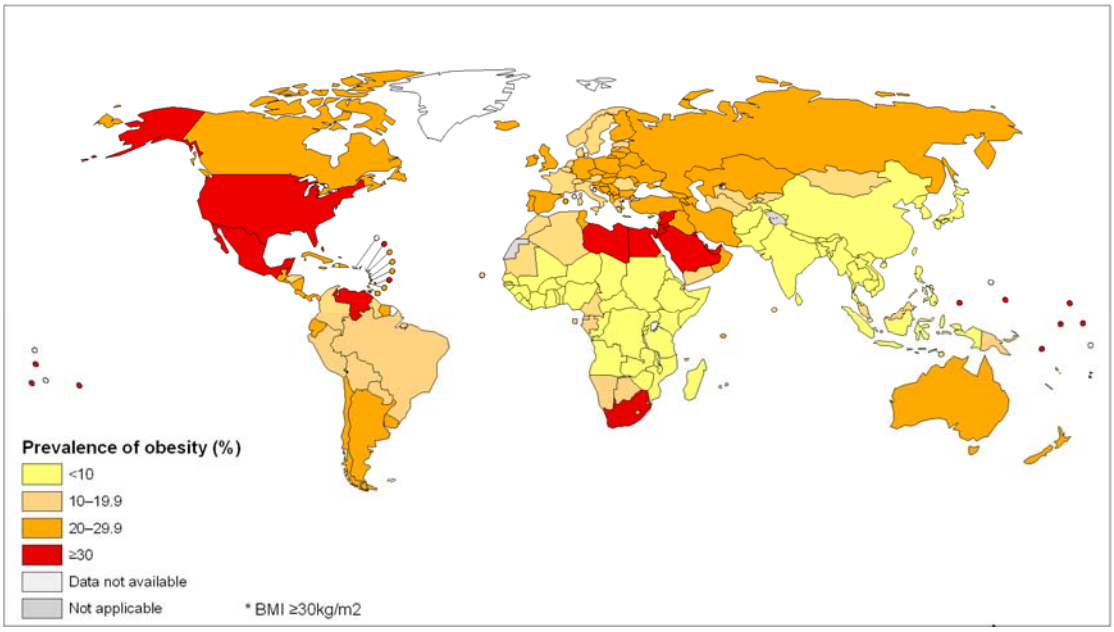

Figure 2. Global prevalence of adult obesity in 2008 (8)**.**

GENETICS OF HUMAN OBESITY

The heritability of obesity has been well-documented, through family (9) and twin studies (10, 11). In these studies, correlation of BMI between identical/monozygotic twins was examined, concluding that environmental differences had little or no influence on BMI. Along with other family studies, heritability for obesity is estimated to be between 40-70% (12), indicating the importance of understanding and elucidating the genes and pathways associated with obesity.

To date, most mutations predisposing humans to severe obesity has been found in the genes of the leptin-melanocortin pathway (Figure 3, Table 1). Clinical reports have described individuals carrying mutations in this pathway to be the cause of an obese phenotype (Table 1). This pathway was first discovered by the identification of leptin as the causative hormone absent in *ob/ob* mice (13).

Leptin is released peripherally by the body's adipocytes upon feeding and binds to its long-form receptor (LepRb) in the arcuate nucleus (ARC), a subregion of the hypothalamus (Figure 3). A population of LepRb neurons in the ARC express proopiomelanocortin (POMC), a 267 amino acid pro-peptide that is a precursor for α-melanocyte-stimulating hormone (α-MSH) (14). The signal is further transduced from the ARC to the paraventricular nucleus (PVN) via the binding of the resulting POMC cleavage product, α-MSH (15). α-MSH binds to the melanocortin 4 receptor (MC4R) and activates the transcription of Single Minded homolog 1 (*Sim1*) and consequently, oxytocin (*OXT*) and other neuropeptides (16, 17). The release of this neuropeptide to the nucleus of the solitary tract (NTS) (18), where it converges with other peptides/hormones, signals satiety and cessation of feeding (19).

One particular member of this pathway, Single Minded homology 1 (*SIM1*), has also been found to be associated with monogenic and syndromic forms of obesity in humans and mice and is the focus of this thesis research.

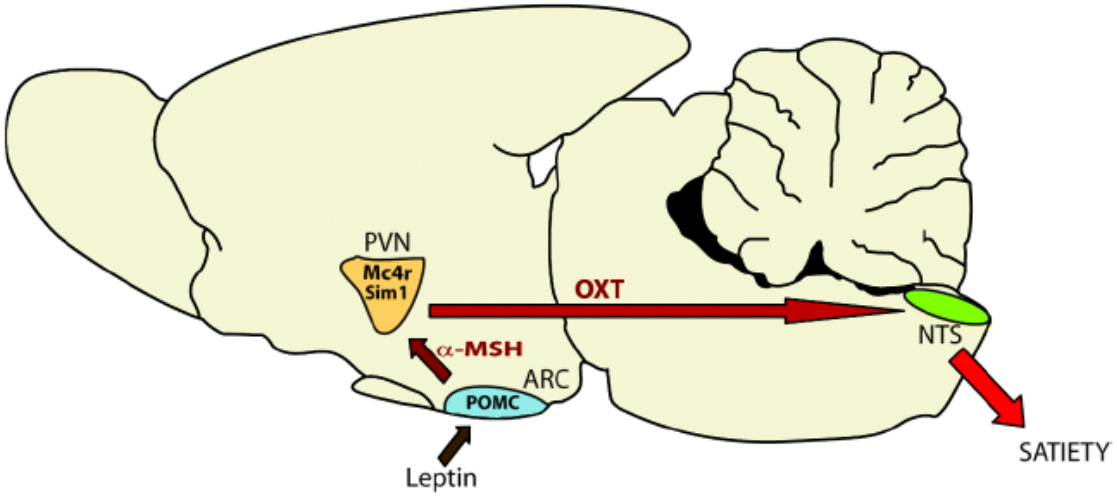

Figure 3. The leptin-melanocortin pathway. Leptin binds to its receptor in the ARC and POMC is synthesized. α-MSH, a cleavage product of POMC binds to the MC4R receptor in the PVN. SIM1 is subsequently activated and OXT signals satiety in the NTS (20).

Gene	Clinical References
LEP	(21–26)
LEPRb	(27, 28)
POMC	(29–32)
MC4R	(33–41)
SIM1	(42–46)

Table 1. Clinical summary of the leptin-melanocortin pathway. Clinical references that highlight gene mutations in the leptin-melanocortin pathway that result in human obesity.

SINGLE MINDED homolog 1 (*SIM1*)

sim was first identified and characterized in *drosophila melanogaster* and was subsequently found to be conserved in vertebrates. The human homologue, *SIM1*, is located on human Chr6q16.3-q21 and its eleven exons span 75,062 base pairs (bp) and encode for a 765 amino acid class E basic helix-loop helix (bHLH) transcription factor (TF) (47). The protein also contains two α-helices linked by a loop that binds to an E-box consensus sequence (CANNTG) (48). In addition, there are two Per-Arnt-Sim (PAS) domains which mediates factor dimerization and stabilization of the DNA binding conformation (49), as well as a nuclear localization signaling (NLS) domain (47) and a domain shared by HIF1α and Trh (HST) proteins (Figure 4).

In drosophila, this gene was shown to be necessary for midline development (50, 51). *sim* mutants fail to derive from a population of midline cells of the neuroepithelium and result in late embryonic lethality due to deficient or incomplete neurogenesis (52). In zebrafish, *sim1* also was shown to be important for the pronephric kidney field (53) and in the developing diencephalon. Knockdown of *sim1* in zebrafish using morpholinos causes significant reduction in isotocin, the zebrafish version of oxytocin and is required for proper isotocin cell development (54).

In mice, *Sim1* is expressed in a variety of tissues, including the fetal kidney and in the central nervous system (CNS) and is essential for the development of the PVN, the supraoptic (SON) and anterior periventricular (aPV) nuclei of the hypothalamus (55–57). Follow-up studies have found that Sim1 heterodimerizes with aryl-hydrocarbon receptor nuclear translocator 2 (Arnt2) and together with Arnt2, Sim1 plays a role in the terminal differentiation of neurons of the developing hypothalamus (58).

Figure 4. SIM1 and its protein domains. The basic helix-loop-helix domain (bHLH) assists in the activation of transcription. Pas-Arnt-Sim1 domains (PAS1 and PAS2) facilitate dimerization and DNA binding stabilization. The HST domain is conserved between HIFα, Sim and Trh proteins and the nuclear localization signaling (NLS) domain interacts with nuclear receptors.

Similar to previous studies that reduce or ablate Sim1 function or expression in drosophila and zebrafish, germline knockout of *Sim1* in mice by two separate groups also exhibited perinatal lethality, due to an underdeveloped hypothalamus in homozygous mice (56, 57). *Sim1* heterozygous mice, are obese, with hyperphagic behavior and increased adipose tissue and have accelerated linear growth (nose to tail length), hyperinsulinemia and hyperleptinemia (55), making these transgenic mice phenotypically identical to *Mc4r*$^{-}$ knockout mice in a dose-dependent manner (15, 59). The major difference between the *Sim1* heterozygous and *Mc4r* full knockout mice was found to be in energy expenditure, which did not seem to contribute significantly, if at all, to the overall increased fat mass and weight phenotype of *Sim1* heterozygotes (55).

Further animal studies place *Sim1* downstream of *Mc4r*, with overexpression of *Sim1* via a BAC transgene partially rescuing the obese and hyperphagic phenotype in the deficient melanocortin signaling agouti yellow A^{y} mice (60). Treatment with melanotan-2 (MTII), an agonist of Mc4r, also demonstrated a blunted response to food intake in *Sim1* heterozygous mice, attributed to the inability to activate *Sim1*-expressing neurons in the PVN (61).

Whether the obesity phenotype is due to the perturbation of the gene during development or to improper activation of the leptin-melanocortin pathway later on in adult time points, was not entirely clear from previous studies. Michaud et al. (55) first asserted that the obesity phenotype was due to hypocellularity of neurons in the hypothalamus in *Sim1* heterozygous mice; however, Holder et al. (56) were not able to confirm this observation in their own independently generated *Sim1* knockout line and have contended that it is the dysfunction of *Sim1*-expressing neurons in regions of the hypothalamus (SON, PVN, basal amygdala, lateral hypothalamic area (LHA)) that causes hyperphagia and obesity.

This question was addressed with the generation of a conditional postnatal knockout of hypothalamic *Sim1*. Both the heterozygotes and full inactivation of the *Sim1* gene led to dose-dependent diet-induced obesity, demonstrating that this gene indeed plays a distinct role in energy homeostasis that did not depend on the proper differentiation of the developing hypothalamus (62). Overexpression of *Sim1* in the PVN of wild-type mice also led to reduced food intake, whereas knockdown by a short hairpin RNA increased food intake, further supporting the hypothesis that postnatal *Sim1* has a distinct function in the PVN to regulate energy consumption (16, 63).

In humans, chromosomal aberrations in the *SIM1* locus result in hyperphagic obesity (42). Individuals who carry interstitial deletions and rearrangements that include the *SIM1* locus have Prader-Willi-like syndrome (PWS-like; OMIM #176270) and are obese (64–70). In addition, sequencing studies on the coding regions of *SIM1* in obese and lean cohorts found an increase in unique rare (minor allele frequency (MAF) <1%) variants in the obese population, second only in prevalence to mutations in *MC4R* (46). A

SIM1 haplotype of two common (MAF<1%) nonsynonymous SNPs (P352T/A371V) in complete linkage disequilibrium have also been identified and associated with obesity (or higher BMI) in severely obese Caucasian males (43) and nominally associated in obese French Europeans (71). Most recently, two groups identified and functionally characterized protein coding mutations in *SIM1*: Bonnefond et al. identified 8 rare variants; 4 variants in morbidly obese individuals with and 4 variants without Prader-Willi-like syndrome features; 7/8 of these protein coding mutations altered *SIM1* activity in functional assays (44). Ramachandrappa et al. discovered 13 rare variants, 9 of which reduced SIM1 function; probands carrying these rare variants segregated with hyperphagic behavior while metabolic rates were normal (45). Furthermore, major allele variants in *SIM1* in the Pima Indian population were also found to be associated with BMI (72). Collectively, this evidence strongly supports *SIM1*'s role in the genetic predisposition to obesity.

As inferred by mouse knockout studies and human clinical reports of *SIM1* deletions, decreased gene dosage or expression levels of *SIM1* can cause obesity, leading to the hypothesis that variations in regulatory elements could, by decreasing the expression of *SIM1*, also be a genetic cause of obesity.

REGULATORY ELEMENTS: ENHANCERS

Regulatory elements include promoters, enhancers, insulators and silencers. For the purpose of this dissertation research project, we have focused on characterizing enhancers. Enhancers are genetic sequences that activate/control the transcription of their target genes (Figure 5) and are important for vertebrate development (73–76). Enhancers

have been found to dictate spatiotemporal gene expression and levels. They are thought to recruit transcription-promoting proteins to the promoter of genes to "enhance" or activate transcription. Studies suggest that enhancers function independent of orientation and/or distance from the target promoter/gene (77). They can be located upstream or downstream or even within the gene that they control, termed as *cis*. Enhancers can also be kilobases (kb)/ megabases (Mb) away or even in *trans*- on a different chromosome from their target gene (78, 79).

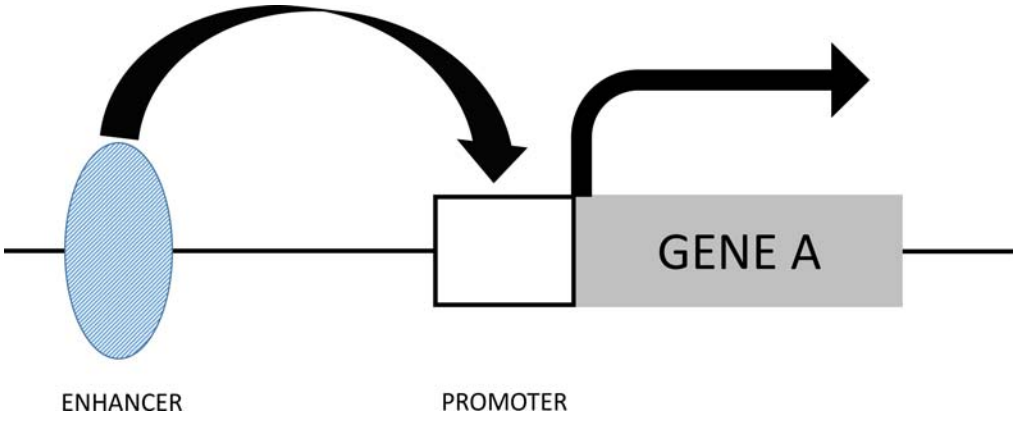

Figure 5. An enhancer regulates transcription via the promoter. A *cis* enhancer (grey oval) is interacting with the Promoter, which leads to the transcription of Gene A, as noted by the arrow above Gene A.

One well known example of a *cis* enhancer is the Sonic Hedgehog (*SHH)* limb enhancer, call the ZPA regulatory sequence (ZRS). This enhancer resides in a neighboring gene, *LMBR1* and is 1Mb away from *SHH* (80). The ZRS is conserved to fish and mutations or insertions/deletions (indels) in the enhancer results in ectopic or altered *Shh* expression in mice, resulting in congenital limb malformations in both humans and mice (81–83).

The H element is example of an enhancer that regulates the expression of olfactory receptor (OR) genes in *trans*. Located on chromosome 14, the H-enhancer element interacts with one of many OR genes on different chromosomes to assist in the fate selection/activation of a single OR allele in an olfactory sensing neuron (84).

In the leptin-melanocortin pathway, enhancers have been characterized for *Pomc*. Two elements approximately 10-12kb upstream of the *Pomc* transcriptional start site (TSS) have been found necessary to regulate its hypothalamic expression (85), while another pituitary-specific enhancer was later found 7kb upstream of the gene (86). Potentially, human variation in these regulatory elements could lead to obesity susceptibility.

COMPARATIVE GENOMICS

With the completion of vertebrate genomes, driven by the advancement of sequencing technologies, genomic alignments and comparisons for sequence similarities between different organisms offers an approach to discover putative enhancer elements. Identifying conserved noncoding (nc) DNA, functional regulatory elements can be detected (73–76, 87–89) based on the hypothesis/principle that these ncDNA sequences

are under purifying selection, similar to coding DNA. The thought is that nonfunctional genetic regions are not under the same constraints and are allowed to become divergent between species due to genetic drift. Hence, enhancers compared to nonfunctional sequences are more conserved – a feature that has been used as a proxy for functionality for gene regulatory elements (90, 91).

FUNCTIONALLY VALIDATING PUTATIVE ENHANCER ELEMENTS

In vivo enhancer assays have been used to validate putative enhancer sequences (74, 75, 93, 94). Various model organisms have been used to validate and characterize enhancer activity, ranging from fly, zebrafish, chicken, frog and mice. The advantages/benefits of these assays are the ability to test the sequence in the context of the whole organism and the real-time assessment of enhancer activity across multiple tissues and time points. Putative enhancer sequences are cloned in front of a minimal promoter (that typically cannot drive expression unless the sequence is an enhancer) followed by a reporter gene (i.e. β–galactosidase (*LacZ*) for mice and Green Fluorescent Protein (*GFP*) for zebrafish) and are used to generate transgenic animals (Figure 6). In this study, I used zebrafish and mouse *in vivo* enhancer assays to assess my enhancer candidates (74, 87, 94–96).

RESULTS

Comparative genomic analysis of the *SIM1* locus

To identify potential *SIM1* enhancers, we carried out a comparative genomic analysis on the *SIM1* locus. We searched for conserved noncoding regions in this locus, defined as one gene upstream (Activating signal cointegrator 1 complex subunit 3 - *ASCC3*) and downstream (Melanin-concentrating hormone receptor 2 - *MCHR2*) of *SIM1*, for a total genomic distance of approximately 1 megabase (Mb) (Figure 7). Analysis of the *SIM1* locus revealed a human-mouse synteny block that ends upstream of *MCHR2* and separates/eliminates any noncoding conservation between human and mouse approximately 93 kilobases (kb) upstream of *MCHR2* (Figure 7). This suggests that enhancers found within the synteny block likely regulate *Sim1* in mice (97). Furthermore, while *MCHR2* is expressed in the hypothalamus in humans and is modestly associated with polygenic obesity (98), *MCHR2* and its ligand, MCH, are not present in mice.

Using the ECR Browser (99), 488 evolutionary conserved sequences (ECRs) between human and mouse were found that were at least 70% conserved for at least 100 base pairs (bp) within the defined *SIM1* locus. ECRs were analyzed manually for repetitive sequences and any RNA coding evidence using the UCSC Genome Browser (100), removing any that contained either. The remaining 360 noncoding ECRs were then ranked by species conservation to prioritize for enhancer assays. Seventeen ECRs, conserved between human and frog, were chosen for subsequent enhancer assays (Table 4) and were termed *SIM1* candidate enhancers (SCEs).

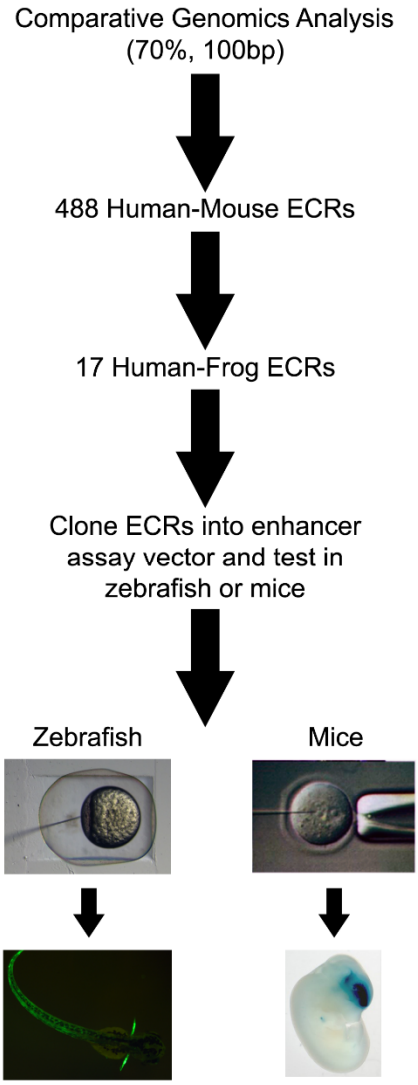

Figure 6. The experimental comparative genomics approach and the zebrafish and mouse enhancer screens. 488 ECRs were found in the *SIM1* locus and 17 human-frog ECRs were chosen for enhancer assays in zebrafish. Tested ECRs were cloned into an enhancer assay vector: an E1B minP-GFP vector (for zebrafish) or an Hsp68 minP-LacZ vector (for mice) and microinjected into one-cell stage *Casper* zebrafish or mice embryos. A 48hpf zebrafish is showing trunk and limb enhancer expression while an E11.5 mouse embryo is exhibiting forebrain enhancer activity, as examples.

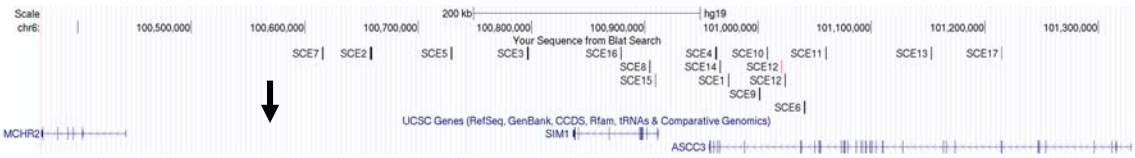

Figure 7. *SIM1* locus and SCEs tested for enhancer activity. UCSC Browser shot of the *SIM1* locus on which comparative genomics analysis was performed and the SCEs tested for enhancer activity. Human-mouse ECRs were identified within one gene upstream (*ASCC3*) and one gene downstream (*MCHR2*) of *SIM1*. The black arrow indicates the end of the human-mouse synteny block that ends approximately 93kb upstream of *MCHR2*.

Zebrafish enhancer screen

The human sequences of the seventeen SCEs were cloned into a vector containing the E1b-minimal promoter (minP) followed by the Green Fluorescent Protein (*GFP*) reporter gene (101). These constructs were microinjected into *Casper* zebrafish embryos at the one-cell stage, using the *Tol2* transposase system for germline integration of the transgene (102). *Casper* zebrafish were utilized for this screen due to their transparency, allowing GFP expression to be easily visualized at both developmental and post developmental stages (103). Enhancer activity was annotated at three developmental time points: 24, 48, 72 hours post fertilization (hpf) and three post developmental time points: 1, 2, 3 months post fertilization (mpf).

From the seventeen sequences, several SCEs demonstrated enhancer activity in the vicinity of the hypothalamus at developmental and post developmental time points as defined by > 20% GFP expression over background: SCE2 and SCE8 during developmental time points and SCE2, SCE4, SCE11 and SCE13 during post developmental time points. In stable zebrafish lines of these six constructs, only SCE2 and SCE8 consistently showed the observed expression patterns across developmental and adult time points. SCE2 and SCE8 exhibited specific GFP expression in the diencephalon/hypothalamus region during both developmental and post developmental time points (Figure 8). The expression of *sim1* at 24-48hpf as characterized by whole-mount *in situ* hybridization (WISH) (54, 104) overlaps the GFP expression of SCE2 and SCE8.

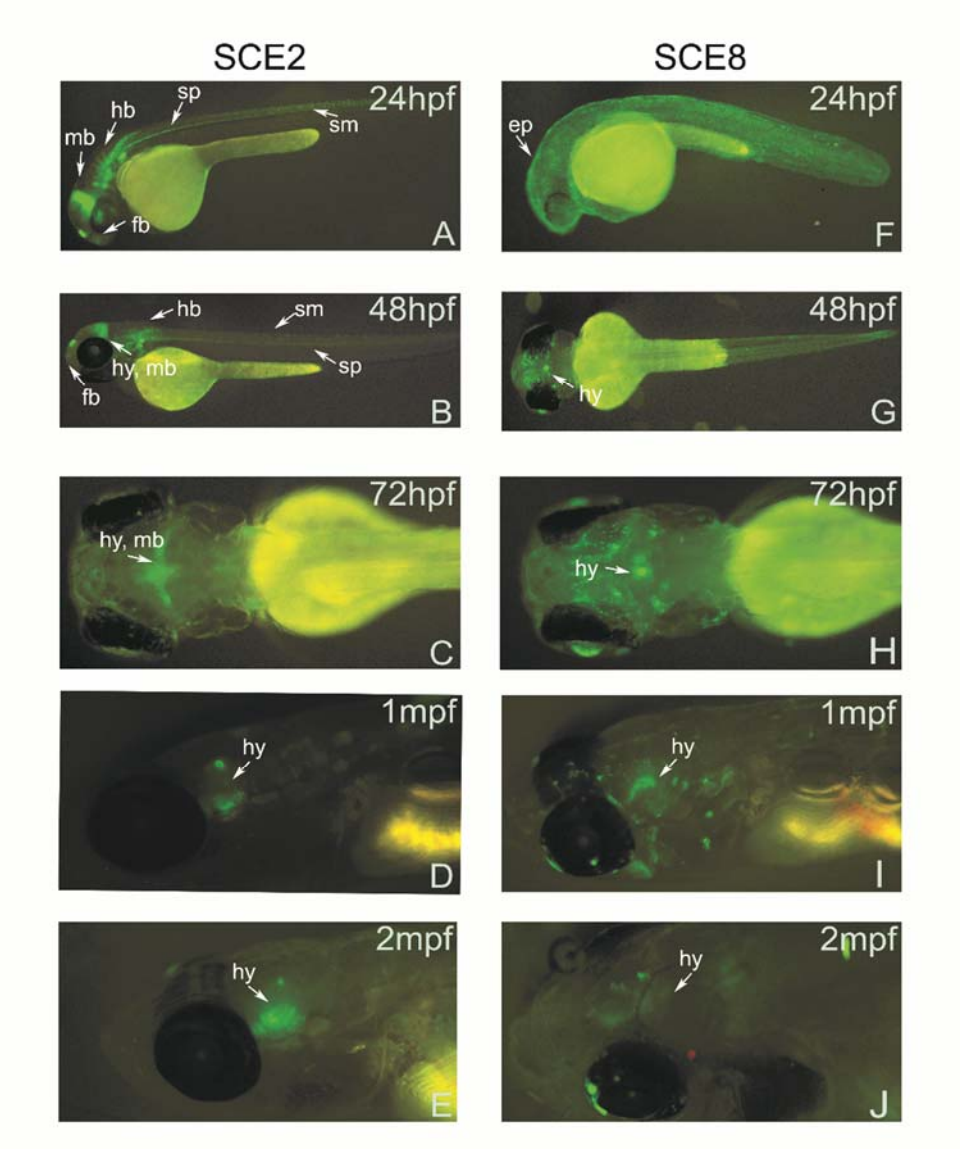

Figure 8. SCE2 and SCE8 act as enhancers across developmental and post developmental time points in zebrafish. (A-D) SCE2 drives expression of GFP in the forebrain (f), hypothalamus (hy), midbrain (m), hindbrain (hb), somites (sm) and spinal cord (sp) during development and localizes to the midbrain/hypothalamic region following development. (E-H) SCE8 is a midbrain and epidermis (ep) enhancer during development and maintains GFP expression in the hypothalamus at later stages.

Mouse enhancer assays

To test whether our zebrafish enhancers have comparable expression patterns in mammals, we tested SCE2 and SCE8 using a mouse transgenic enhancer assay (75). SCE2 and SCE8 were cloned into the Hsp68-LacZ reporter plasmid (105) and were used to generate transgenic mouse embryos using standard techniques (106). SCE8 showed enhancer activity in the developing hypothalamus and midbrain and overlapped *Sim1* expression at E12.5 (Figure 9, Figure 10), a time point during hypothalamic development at which *Sim1* expression has been documented at high levels (107–109). SCE2 was previously found to be an enhancer in the developing diencephalon, mesencephalon and cranial nerve at E11.5 (110).

Since SCE2 showed the strongest, consistent expression in the developing hypothalamus both in zebrafish and mice, we further investigated its enhancer activity in mice across developmental and adult time points. A stable SCE2 transgenic mouse line was generated and LacZ expression was assayed at different time points. In addition to repeating previous findings at E11.5, we found that the activity of SCE2, as reflected by LacZ staining, is evident at E9.5, maintained during adulthood and is similar to that of *Sim1* (Figure 11). Enhancer activity is localized primarily throughout the developing hypothalamus up to E12.5 and transitions from that region to the hippocampus from E13.5 to E15.5 (Figure 12). Postnatally, enhancer activity is observed in the hypothalamus and remains in the hippocampus at P56, similar to *in situ* hybridization on mouse brain sections for *Sim1* (Figure 11) (111).

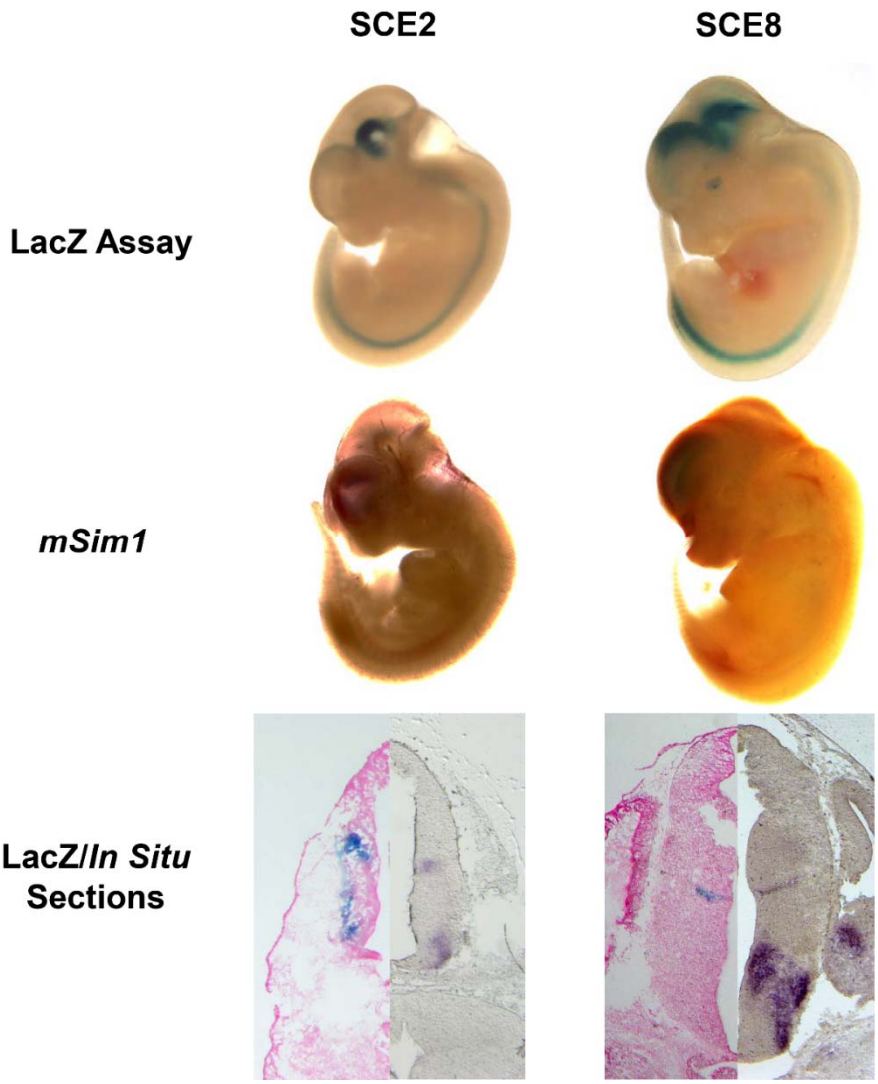

Figure 9. SCE2 and SCE8 are mouse enhancers during development. SCE2 is active in the diencephalon, midbrain and neural tube at E11.5 and recapitulates *Sim1* expression at this time point. SCE8 shows enhancer activity in the developing hypothalamus at E12.5 that overlaps *Sim1* expression in the P2-P3 regions of the diencephalon. Sections on the bottom show SCE 2 and SCE8 expression on the left compared to *Sim1* expression, as determined by *in situ* hybridization at E11.5 and E12.5 respectively.

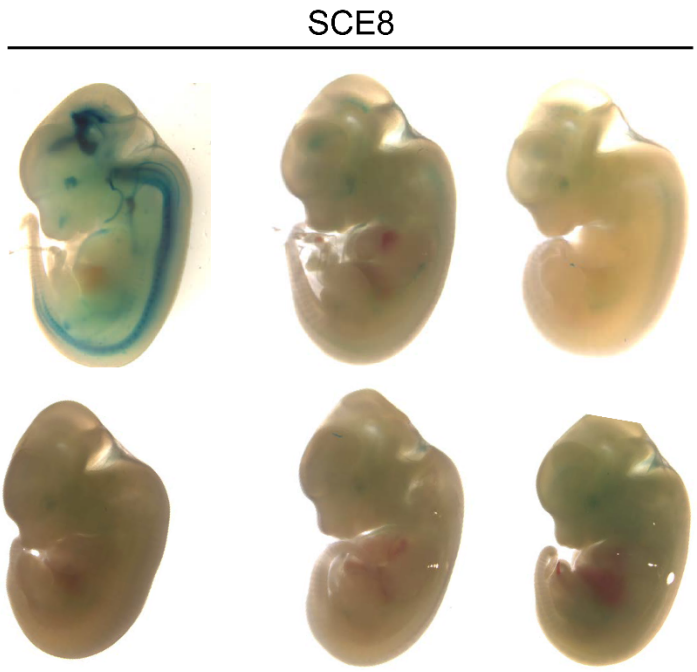

Figure 10. SCE8 LacZ positive transgenic mice (E12.5). Three out of six mice exhibited consistent diencephalic and midbrain enhancer activity as reflected by LacZ staining.

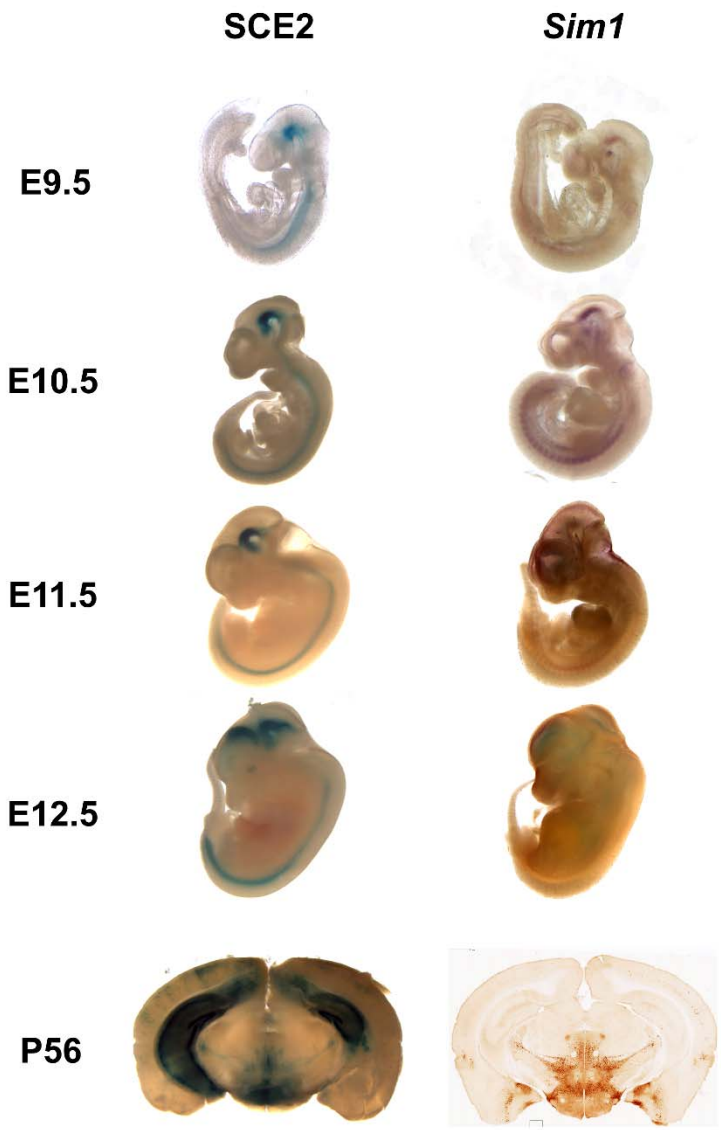

Figure 11. SCE2 enhancer activity recapitulates *Sim1* expression. SCE2 enhancer activity across various time points as reflected by LacZ staining. SCE2 enhancer activity overlaps *Sim1* expression from E9.5 to E12.5 in the diencephalon (developing hypothalamus) and mesencephalon (midbrain) and in the adult hypothalamus. Adult hypothalamus *in situ* expression was obtained from GENSAT (111)

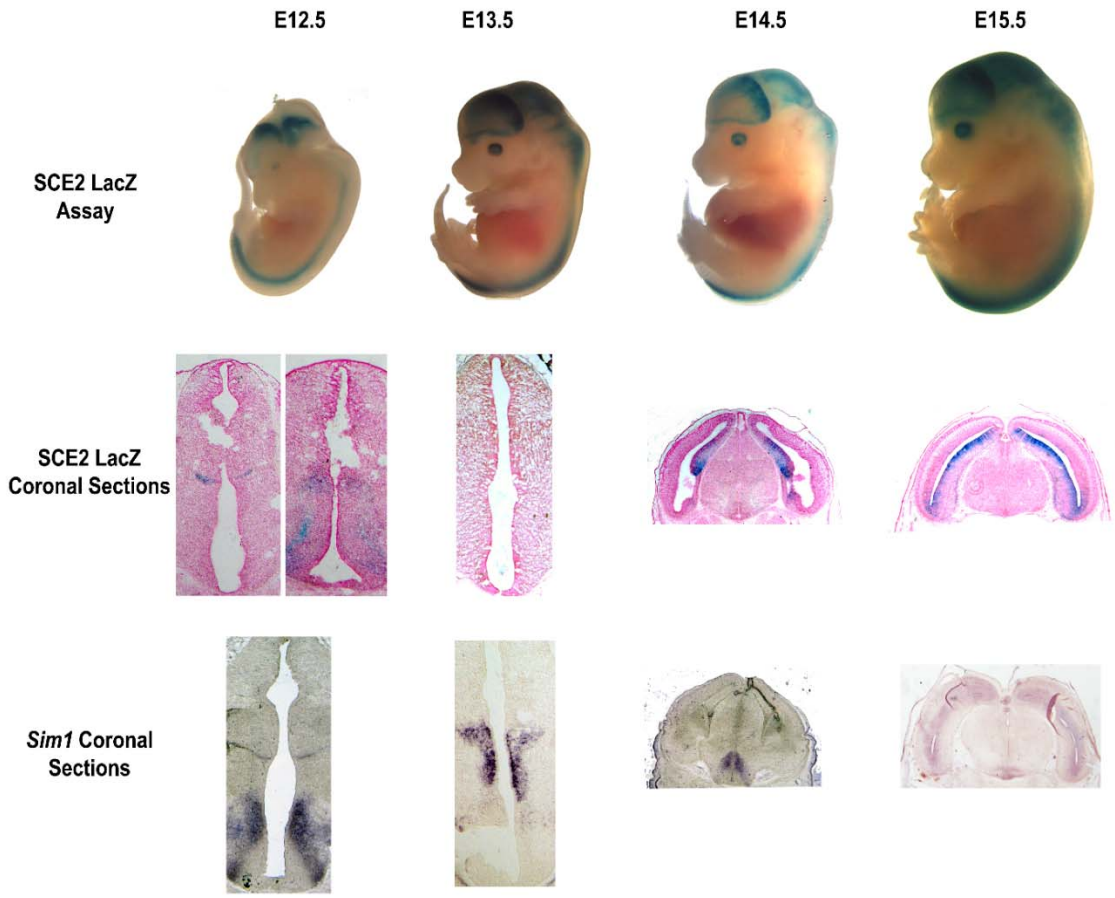

Figure 12. SCE2 is active during development but does not overlap Sim1 expression from E13.5-15.5. SCE8 E12.5 enhancer assay midbrain sections compared to *Sim1* expression in the diencephalon (20 microns).

SCE2 deletion series identifies a functional hypothalamus enhancer core

To further refine the domain driving hypothalamus expression in SCE2, we carried out a deletion series in zebrafish. SCE2 is 87% conserved between human and mouse across the length of the enhancer (Figure 13A). Further genomic comparisons reveal an approximately 700bp sequence, named here as the core element (CE), that is conserved between human and fish. In addition, examination of a mouse E11.5 forebrain E1A binding protein p300 (EP300 or p300; a protein that co-localizes with enhancers) in a chromatin immunoprecipitation followed by sequencing (ChIP-seq) dataset (112) identified a peak that overlaps the CE. Combined, the evolutionary conservation and p300 ChIP-seq peak suggested that this 700bp CE within SCE2 could be the major functional domain of this enhancer and that it may be sufficient to drive the enhancer expression observed in our original screen.

The CE and the flanking sequences (5' and 3' elements relative to *SIM1*) that comprise SCE2 were individually cloned into the E1B-GFP zebrafish enhancer vector and microinjected into zebrafish as previously described. The CE was sufficient to drive the majority of SCE2 *GFP* expression in the forebrain, midbrain, hindbrain and somites at 24-72 hpf (Figure 13C). The 3' element appeared to drive the spinal cord expression pattern, the only pattern that was not observed in the enhancer assay for SCE2-CE (Figure 13B). The 5' element was negative for GFP expression (Figure 13D). These results suggest that the CE of SCE2 is the functional domain that is important for the hypothalamus expression of this enhancer.

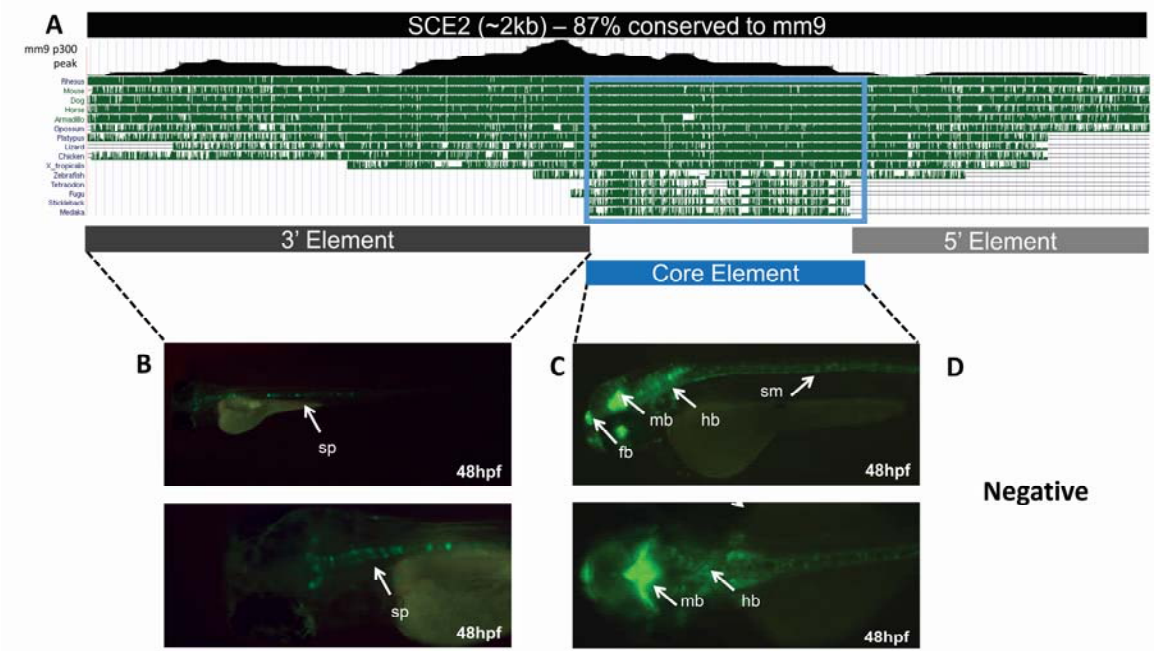

Figure 13. SCE2 enhancer deletion series. (A) SCE2 is 87% conserved between human and mouse and has a 700bp core sequence that is conserved between human and fish. (B) The 3' region of SCE2 drives GFP expression in the spinal cord (sp). (C) The core element (CE) drives enhancer activity in the forebrain (fb), hypothalamus (hy), midbrain (mb) hindbrain (hb) and somites (sm). (D) The 5' SCE2 sequence was negative for enhancer activity.

To verify this in mammals, we also generated a transient transgenic of SCE2-CE in mice. The construct was cloned into the Hsp68LacZ vector and microinjected as previously described. The LacZ enhancer assay at E11.5 demonstrated that SCE2-CE reflects the original enhancer pattern observed in the fuller insert, however, we observed that LacZ expression was more ectopic in comparison to full sequence version of SCE2 (Figure 14).

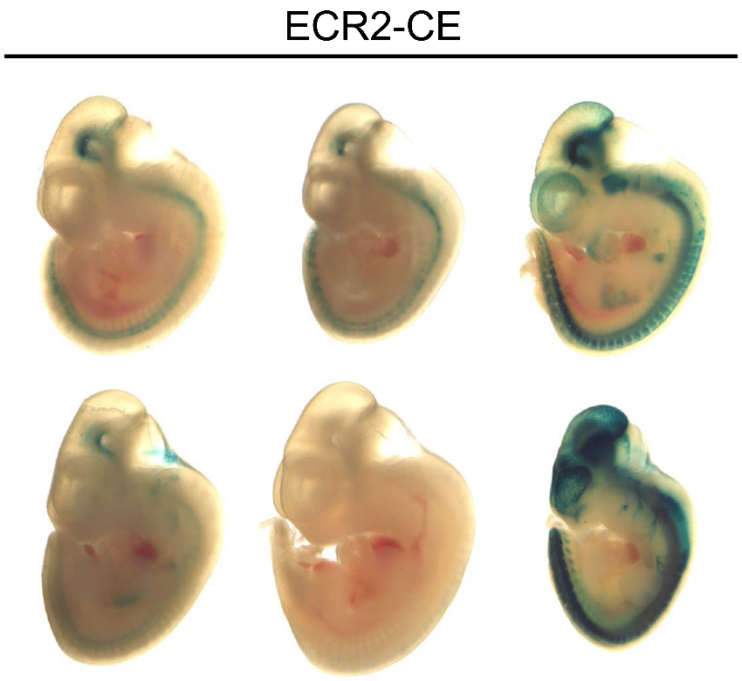

Figure 14. SCE2-CE exhibits similar enhancer activity compared to the full length SCE2. LacZ enhancer assay for SCE2-CE was performed on E11.5 mice. 5/6 mice recapitulated SCE2's enhancer pattern, however, the pattern is more ectopic and nonspecific than SCE2.

Sequencing analysis of SCE2-CE in obese and lean cohorts

To determine whether nucleotide variants in this enhancer could be associated with predisposition to obesity, we sequenced a large obese cohort (N=510) and compared our results to a matched lean cohort (N=554) (37). We focused on the SCE2-CE region, since our zebrafish deletion series analysis demonstrated that this sequence is responsible for hypothalamus enhancer activity. Two common SNPs: rs187302227 and rs192532320, which flank the portion of SCE2-CE that is conserved to fish, were present in both cohorts and did not show any significant MAF frequency differences (Table 2). However, four novel rare variants that were unique to the obese cohort were found as singletons, one of which (chr6:100658719 G>A, was found in a homozygous form) versus one unique variant (rs182500930) in the lean cohort (Table 2).

To provide further support that the rare variants present in the obese individuals are unique, we analyzed the prevalence of variants in SCE2-CE in the 1000 Genomes Project (113). Other than the two flanking common SNPs (rs187302227, rs192532320) previously found in the obese and lean cohorts, five other rare variants were identified in SCE2-CE in the 1000 Genomes Project (Table 3). Once weighted according to the race/ethnicity percentages of our obese cohort (85% White, 10% Black, and 5% Latino race/ethnicity), two variants, rs150264974 and rs188023826, were present with weighted MAFs of 0.0008 and 0.0002, respectively. These variants were not shared by our case or control groups, suggesting that the variants identified in our obese cohort are unique to the obese cohort. In summary, these results show a potentially higher prevalence of rare variants in the obese versus the lean cohort and the 1000 Genomes Project.

SCE2-CE Variant Chromosomal Position (hg19)	dbSNP #	Base Change	Conservation	MAF (obese)	MAF (lean)
chr6:100658489	rs187302277	T/C	mixed	0.0114	0.0049
chr6:100658548	NA	C/T	mammals	0.0014	
chr6:100658624	NA	C/T	stickleback	0.0014	
chr6:100658893	rs182500930	C/T	stickleback		0.0012
chr6:100658719	NA	G/A	stickleback	0.0014	
chr6:100658735	NA	G/A	stickleback	0.0028	
chr6:100659108	rs192532320	C/A	opossum	0.0055	0.0049

Table 2. Sequencing summary of SCE2-CE. Sequencing of an obese cohort identified four rare SNPs that are not present in the lean controls cohort. One SNP (rs182500930) unique to the lean group is not present in the obese cohort and was identified in dbSNP and the two flanking SNPs (rs18730227, rs192532320) around the conserved area have also been previously reported and present in both sequencing cohorts without significant MAF differences.

				MAFs				
CHR	SNP	A1	A2	White	Black	Asian	Latino	White85%-Black10%-Latino5%
6	rs187302277	C	T	0.020	0	0	0.0030	0.017
6	rs190692772	C	T	0	0	0.0017	0	0
6	rs182500930	G	T	0	0	0.010	0	0
6	rs150264974	C	T	0	0.0077	0	0	0.00077
6	rs145891113	G	T	0	0	0.0017	0	0
6	rs188023826	T	C	0	0	0	0.0030	0.00015
6	rs192532320	A	C	0	0	0.0033	0.0053	0.00027

Table 3. SCE2-CE 1000 Genome SNPs. Seven SNPs were identified in the SCE2-CE region, including the two previously identified SNPs from the obese and lean cohorts: rs187302277, rs182532320. However, stratification of individuals in the 1000 Genomes Project by race/ethnicity that reflects the case and control cohorts results in one SNP that is present within SCE2-CE: rs150264974.

Functional characterization of obese-specific SCE2-CE variants

The four obese-associated variants were functionally analyzed for differential enhancer activity during development using the zebrafish enhancer assay. Using site-directed mutagenesis, we cloned all four variants into SCE2-CE. The variants were microinjected into zebrafish embryos, as previously described, and annotated at developmental time points (24, 48 and 72 hpf). We observed that, in general, the original pattern of SCE2-CE was maintained during developmental time points, regardless of the variant (data not shown). However, GFP-expressing fish showing a similar enhancer pattern differed between the variants compared to the reference sequence. The most notable difference was a statistically significant reduction in hypothalamic enhancer activity for SCE2-CE-100658719 G>A across all three time points (Figure 15). Combined, these results suggest that this variant could alter the enhancer activity of SCE2 on *SIM1*.

Transcription factor binding site analyses of the four obese-segregating and one lean variants were performed. TRANSFAC analysis on the obesity-specific rare variants revealed that only the SCE2-CE-100658719 G>A variant was predicted to result in a gain of a CCAAT-enhancer- binding protein (C/EBP) binding site; this is corroborated by Genomatix and rVISTA TFBS analyses of the variant. The other variants did not result in a loss or gain of TFBS.

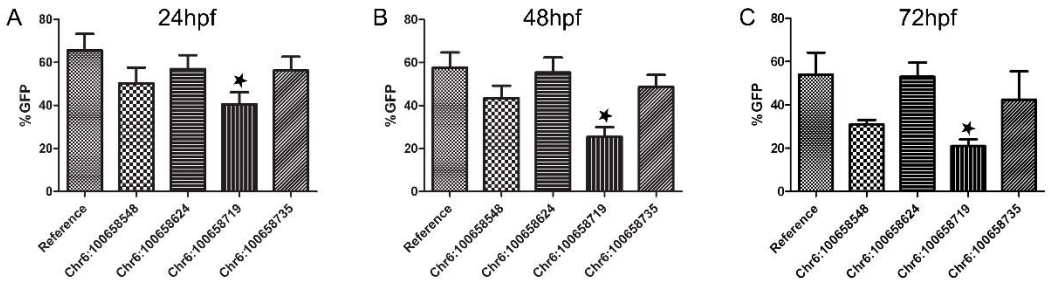

Figure 15. Zebrafish hypothalamic enhancer activity of obesity-associated SCE2-CE variants. Variant 100658719 exhibited statistically significant reduction of hypothalamic enhancer activity across developmental time points ($p<0.05$; One-way ANOVA followed by Dunnett's multiple comparison test).

Discussion

Using comparative genomics coupled with *in vivo* enhancer assays, this study identified two novel midbrain enhancers in the *SIM1* locus. Both enhancers showed similar hypothalamic activity in zebrafish and mice at comparable developmental time points, further demonstrating the validity of using zebrafish as a rapid and cost efficient filter to detect mammalian enhancers in this tissue. Furthermore, zebrafish were used to identify the core region within SCE2 that drives enhancer activity in the hypothalamus. Sequencing SCE2-CE in an obese cohort identified four unique rare variants. Functional characterization of these variants in zebrafish showed that one variant, chr6:100658719 G>A, exhibited reduced enhancer activity, with statistically significant enhancer activity reduction during development.

In this study, we used comparative genomics to identify enhancers in the *SIM1* locus. Although we identified enhancers that may regulate *SIM1*, there could well be additional enhancers in this locus that control the spatiotemporal expression of the gene. The use of techniques like chromatin immunoprecipitation followed by sequencing (ChIP-seq) on specific tissues (e.g. mouse hypothalamus) using a variety of enhancer marks, might help identify additional tissue-specific enhancers in this locus. Currently, no hypothalamus-specific ChIP-seq datasets have been published, especially in the context of energy homeostasis and the signaling cascade of which *SIM1* is a member. Such future assays would help better characterize the regulatory landscape of this region.

The two enhancers we identified were active in the developing diencephalon during neurogenesis. SCE8 is active in both the PVN and a small region of the mammillary epithelium, both of which express *Sim1* at E12.5 (Figure 9, Figure 10) (109). Spatiotemporal investigation of the SCE2 enhancer shows that it is active throughout

development and maintains hypothalamic enhancer activity in adult mice. SCE2 activity overlaps *Sim1* expression in the earlier stages of neurogenesis (E9.5-E12.5) (Figure 9; Figure 11) and in the adult mouse hypothalamus (Figure 11). However, SCE2 was found to be active in the hippocampus from E13.5 to adulthood (Figure 11, Figure 12) where *Sim1* is not known to be expressed (111). This pattern of expression has been previously observed for the Pro-opiomelanocortin (*Pomc*)-*GFP* mouse model, a gene in the leptin-melanocortin pathway, but is primarily used to mark granule cells in the dentate gyrus/hippocampus (114). It is worth noting that this expression pattern could also be due to the site of integration of the transgene, being influenced by enhancers in that region (though it was observed in both founders) or background from the minimal promoter. Further studies, such as a mouse knockout of SCE2 or chromosome conformation capture, would be needed to determine whether SCE2 truly interacts and transcriptionally controls *Sim1* in either the developmental or energy homeostasis context.

In comparison to SCE2, SCE8 is also active in the developing PVN and mammillary epithelium (Figure 9), suggesting there is redundancy in the potential regulation of *Sim1* expression at E12.5 and perhaps during other periods of development. This could be attributed to a phenomena termed “shadow enhancers”, where another enhancer could have a similar expression pattern so as to provide backup for the other enhancer (115).

Sequencing of SCE2-CE identified novel rare variants that were unique to the obese cohort and were not observed in either the lean cohort or in any of the populations analyzed from the 1000 Genomes Project (Table 3). While the difference in prevalence of variants may not be significant between populations, the uniqueness of the SNPs

identified in the obese cohort supports previous studies that report a higher prevalence of unique *SIM1* variants in individuals with severe morbid obesity (44–46). The observed changes in enhancer activity for SCE2-CE-100658719 G>A in zebrafish (Figure 15) suggest that this variant could alter *SIM1* expression. However, this assay has an important limitation that needs to be taken into account. It is difficult to make quantitative conclusions due to the *Tol2* transposase system allowing for variable integration sites and copies of the transgene into the zebrafish genome (116). While we tried to correct for this, by using an increased sample size (>100 eggs/per each injected variant), this caveat and poor survival rates of microinjected *Casper* embryos that can additionally skew numbers and needs to be kept in mind regarding our results.

Ultimately, associating functional enhancer variants with obesity could further elucidate the genetic contributions to this phenotype. While coding exons have been the focus of obesity genetics, it is important to acknowledge that regulatory elements could also contribute to the genetic predisposition of human disease and phenotypes. With technological developments such as ChIP-seq (117), massively parallel reporter assays (118), whole-genome sequencing (WGS) and other robust applications of advanced sequencing technologies, we will be able to further elucidate regulatory genetic elements that are associated with human variation and disease. Overall, these approaches will contribute to our ability to investigate genetic variants that segregate with currently unclear phenotypes. However, we need to keep in mind that even with the identification of rare variants from WGS of individuals with rare and undiagnosed phenotypes, it is still difficult to diagnostically conclude an identified variant, mutation or SNP as the casual nucleotide for a particular/given phenotype. While there may be one variant associated

with a gene that is a reasonably responsible for an individual's phenotype, the numbers of variants identified by WGS for each individual gives us thousands of variants to consider before determining the casual variant(s). A combinatorial approach such as was done in this thesis and data generated by ChIP-seq, WGS and other high throughput-driven studies will help investigate and elucidate these genetic variants and their contribution to human disease.

MATERIALS & METHODS

Comparative genomics

Using ECR Browser (99), we selected intronic and intergenic ECRs between *MCHR2* and *ASSC3* that were ≥100bp long with at least 70% sequence identity between human and mouse. This analysis generated 488 unique ECRs. We manually filtered out repetitive sequences, expressed sequence tags (ESTs) and any other coding sequences using the UCSC Genome Browser (100). To prioritize ECRs for functional assays, the remaining 360 ECRs were then ranked by species conservation. A total of seventeen human and frog ECRs (labeled as SCEs) were chosen for enhancer assays (Table 4).

ECR	Chromosomal Location (hg19)	Length (bp)	Primers
SCE1	chr6:100973720-100974484	769	GATGCATTTAGCTTTCCCAAAAC
			TTGATCCTAGGCAGCTTGTTAAA
SCE2	chr6:100657651-100659532	1886	GACCATTGGTGAGGGAAGATTA
			TTTCCGTAACATAACACCCAAAT
SCE3	chr6:100796115-100797773	1663	GGCTGTTTTCCTTCTGTTAATTG
			TAAAGAAAAGACTGCCTTGGCTA
SCE4	chr6:100962273-100963587	1317	CCCCTCTCTTATTATAATCACTTTGC
			TAACATTTACCCCATGAAGCAAG
SCE5	chr6:100728914-100730115	1206	AAATTCGCATTGGATTTTACAGA
			ATCAAAACAGAAAGTGGTGATGG
SCE6	chr6:101040567-101041736	1174	TGGATTCAATTTGTATGTTCAGTT
			TTCTAGCCTGAACAATTATAAGCA
SCE7	chr6:100615387-100616553	1170	CTGTATTTCCTCAATGCTTCCAC
			TTTCCTTGAGTCTCTCAAACTCATT
SCE8	chr6:100904128-100905067	943	CTTCCTCCTCTTCCTGTCAGAGT
			AAGTGGGGTAGAGACCTTCTCC
SCE9	chr6:101000766-101001774	1012	CAGCATGAATATGCAAGCAAAA
			TTTCTTTTCCTTTTGGAGATGAT
SCE10	chr6:101007935-101008614	682	CCTGTCAATCAAATAGCTAATGC
			TGTTTGGGGTTTGACTATGACTT
SCE11	chr6:101059749-101060409	664	CATCTTTTATTGAGCATTTCAGACA
			AAAAACCATGATAAATGAAACACTGA
SCE12	chr6:101023754-101024381	630	CCCTATTTCCTTTTCGACATATTCT
			ACCTTCTAAAAGTCCACTGAACA
SCE13	chr6:101152660-101153281	624	TTCAATTTGTTTCTCCTTATTTGG
			TGAAGGTAGATTTTCAGTAAGAAGC
SCE14	chr6:100966364-100966863	503	CATTAGCAGCTGAATTTAAAGAGTGA
			CATTTTAATTATTTGCTAGGCATATCT
SCE15	chr6:100909236-100909685	454	GGGGAACTTATATACTAAAAACCACA
			CCCGGGTCTACACCAAGC
SCE16	chr6:100878939-100879406	472	TGTCTGAATACTGAAGCACAGAA
			CAACACAAAGACACTGGCATTTC
SCE17	chr6:101214612-101215004	397	GACAACAGCTTTTTAAATTGTAAGG
			GAGCAAAAACTACCATGTCATTTG

Table 4. List of the SCEs selected for zebrafish enhancer assays.

PCR amplification and cloning

Candidate sequences were PCR amplified from human genomic DNA (Roche) using TopTaq (Qiagen). Primers were designed to have an additional 100- 200bp flanking the ECR sequence (Table 4); previous experiments have shown this to be a reliable method for obtaining positive enhancer activity (75). Inserts were first cloned into the pENTR-dTOPO vector (Life Technologies) following the manufacturer's protocol and then transferred using Gateway technology (Life Technologies) into the E1b-GFP-Tol2 (101) for the zebrafish enhancer assays and into the Hsp68-LacZ vector (105) for mouse enhancer assays. Orientation and sequence of the inserts were verified by restriction enzyme digest and sequencing. Plasmid DNA was generated for microinjections using the EndoFree Plasmid Midi Prep kit (Qiagen).

Enhancer assays

For zebrafish enhancer assays, each construct was injected into *Casper* (103) embryos at the one-cell stage, following standard procedures (119, 120), along with *Tol2* mRNA (116) to facilitate genomic integration. A minimum of 100 embryos per construct were annotated for GFP expression at 24, 48 and 72 hpf and at least 12 fish were annotated at 1, 2 and 3 mpf. An enhancer was considered positive if 20% of the GFP expressing fish showed a consistent expression pattern after subtracting tissue expression pattern percentages of the negative control (E1b-GFP empty vector) at all respective time points. To generate stable lines, embryos that exhibited GFP expression were selected to mature to adulthood. Once mature, zebrafish were backcrossed to *Casper* zebrafish; this allowed

for the germline transmission of the enhancer-driven GFP pattern. At least two founders were used for the generation of stable lines in zebrafish.

For the mouse enhancer assays, transgenic mice were generated by Cyagen Biosciences using standard procedures (106). For the SCE2 stable line, two founders were generated and analyzed in all time points. At embryonic time points, embryos were harvested and stained for LacZ using standard procedures(75). For the adult time point, P56, LacZ enhancer assays were performed on mouse brains using previously described procedures (121). All animal work was approved by the UCSF Institutional Animal Care and Use Committee.

In situ hybridization

The mouse *Sim1* vector (107) was graciously given to us by the Rubenstein lab and used as a template to generate a digoxygenin-labeled probe. The *Sim1* vector was digested with XhoI and T7 polymerase (Roche) was used to generate the probe (750 bp); an illustra MicroSpin G-50 column (GE Healthcare) was utilized to purify the probe before quantifying. Embryonic mice were harvested and fixed in 4% PFA and *Sim1 in situ* hybridization assays were performed on whole mount and 20 micron cryosectioned mouse embryos (E11.5-E15.5) according to standard protocols (122, 123).

SCE2 enhancer deletion series

Primers were designed around the SCE2-CE region and the flanking regions of the CE (Table 5). The three regions were cloned into the E1b-GFP vector and microinjected into

Casper zebrafish embryos as previously described. Zebrafish embryos were annotated for GFP expression at three developmental time points (24, 48 and 72hpf).

Construct	Chromosomal Location (hg19)	length (bp)	Primer
SCE2-5'	chr6:100659051-100659182	132	F: CCCTAGAGATGTTTGGATAAGTCAA
			R: AACAACCTTAAAATCCTTCTTCAG
SCE2-CE	chr6:100658425-100659182	758	F: AACAACCTTAAAATCCTTCTTCAG
			R: CGGCCACCCTGTCTTTAGT
SCE2-3'	chr6:100657653-100658521	869	F: GTGGTCTGGGTGATCTCAT
			R: TCCGTAACATAACACCCAAATG

Table 5. Primers used for SCE2 deletion analysis.

SCE2-CE sequencing analysis

Sequencing of SCE2-CE was performed on previously reported obese and lean cohorts (37). Briefly, severely obese individuals (n = 510) had a mean BMI of 47.9 $\pm$ 8.3kg/m2, age 48.3 +12.1 years, 73% female and 85% Caucasian. Controls (n = 554) had an average BMI of 22.9 $\pm$ 1.4kg/m2, and were age (51.3 +4.5 years), sex (68% female) and ethnically (82% Caucasian) matched. All work was conducted under approved protocols from the UCSF Committee on Human Research, and informed written consent was obtained from all patients.

Estimated allele frequencies for the SNPs in the SCE2-CE were calculated from the 1000 Genomes Project, March 2012 interim release (Table 3) (113). To reflect the ancestry of our population, we calculated them as a weighted average of 85% European ancestry (87 CEU: Utah residents with ancestry from Northern and Western Europe from Centre, 93

FIN: Finnish from Finland, 89 GBR: British individuals from England and Scotland, 14 IBS: Iberians in Spain, 98 TSI: Toscani in Italia), 10% African ancestry (97 LWK: Luhya in Webuye Kenya, 88 YRI: Yoruba in Ibadan), and 5% Latino race/ethnicity (60 CLM: Columbian in Medellin, Columbia, 66 MXL: Mexican in Los Angeles, 55 PUR: Puerto Rican in Puerto Rico).

Differential enhancer activity of obesity-segregating SNPs

Unique SNPs identified in the obese cohort were generated by site directed mutagenesis of the E1b-GFP plasmid containing the SCE2-CE insert using QuikChangeII (Agilent Technologies). SCE2-CE variants were purified and microinjected into one-cell *Casper* embryos as described above. The enhancer activity of each of the variants was annotated at three developmental time points (24hpf, 48hpf, 74hpf); at least 100 surviving embryos were annotated at each time point. One-way ANOVA followed by Dunnett's multiple comparison test was performed on the variants' percent GFP compared to the reference sequence for statistical analysis using Prism 5.04 (GraphPad).

Transcription factor binding site analysis of obesity-segregating SNPs

The five conserved SNPS identified in the obese and lean cohorts via sequencing was also analyzed for changes in TFBS. Each SNP was examined for loss or gain of TFBS by TRANSFAC (124) using standard default parameters. Results were also validated by Genomatix (125) and rVISTA (126) under their respective standard parameters.

REFERENCES

1. National Center for Health Statistics (2012) NCHS Data on Obesity.

2. World Health Organization (2013) Overweight: Situation and trends.

3. World Health Organization (2013) Overweight and obesity. *Overweight and Obesity*.

4. Haslam,D.W. and James,W.P.T. (2005) Obesity. *Lancet*, **366**, 1197–1209.

5. McTigue,K.M., Hess,R. and Ziouras,J. (2006) Obesity in older adults: a systematic review of the evidence for diagnosis and treatment. *Obesity*, **14**, 1485–1497.

6. Guh,D.P., Zhang,W., Bansback,N., Amarsi,Z., Birmingham,C.L. and Anis,A.H. (2009) The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health*, **9**, 88.

7. CDC (2011) Prevalence of Self-Reported Obesity Among U. S. Adults.

8. World Health Organization (2008) Prevalence of obesity.

9. Comuzzie,A., Higgins,P., Voruganti,S. and Cole,S. (2010) Cutting the fat: the genetic dissection of body weight. *Progress in Molecular Biology and Translational Science*, **94**, 194–212.

10. Stunkard,A. and Harris,J. (1990) The body-mass index of twins who have been reared apart. *New England Journal ...*, **322**, 1483–1487.

11. Maes,H., Neale,M. and Eaves,L. (1997) Genetic and environmental factors in relative body weight and human adiposity. *Behavioral Genetics*, **27**, 325–351.

12. Barsh,G.S., Farooqi,I.S. and O'Rahilly,S. (2000) Genetics of body-weight regulation. *Nature*, **404**, 644–651.

13. Zhang,Y., Proenca,R., Maffei,M., Barone,M., Leopold,L. and Friedman,J.M. (1994) Positional cloning of the mouse obese gene and its human homologue. *Nature*, **372**, 425–432.

14. Schwartz,M.W. and Porte,D. (2005) Diabetes, obesity, and the brain. *Science (New York, N.Y.)*, **307**, 375–379.

15. Huszar,D., Lynch,C. a, Fairchild-Huntress,V., Dunmore,J.H., Fang,Q., Berkemeier,L.R., Gu,W., Kesterson,R. a, Boston,B. a, Cone,R.D., et al. (1997) Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell*, **88**, 131–141.

16. Xi,D., Gandhi,N., Lai,M. and Kublaoui,B.M. (2012) Ablation of Sim1 neurons causes obesity through hyperphagia and reduced energy expenditure. *PLoS one*, **7**, e36453.

17. Duplan,S.M., Boucher,F., Alexandrov,L. and Michaud,J.L. (2009) Impact of Sim1 gene dosage on the development of the paraventricular and supraoptic nuclei of the hypothalamus. *The European Journal of NeuroscienceJ*, **30**, 2239–2249.

18. Peters,J.H., McDougall,S.J., Kellett,D.O., Jordan,D., Llewellyn-Smith,I.J. and Andresen,M.C. (2008) Oxytocin enhances cranial visceral afferent synaptic transmission to the solitary tract nucleus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **28**, 11731–11740.

19. Perello,M. and Raingo,J. (2013) Leptin activates oxytocin neurons of the hypothalamic paraventricular nucleus in both control and diet-induced obese rodents. *PloS one*, **8**, e59625.

20. Kublaoui,B.M., Gemelli,T., Tolson,K.P., Wang,Y. and Zinn,A.R. (2008) Oxytocin deficiency mediates hyperphagic obesity of Sim1 haploinsufficient mice. *Molecular Endocrinology*, **22**, 1723–1734.

21. Montague,C.T., Farooqi,I.S., Whitehead,J.P., Soos,M. a, Rau,H., Wareham,N.J., Sewter,C.P., Digby,J.E., Mohammed,S.N., Hurst,J. a, et al. (1997) Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*, **387**, 903–908.

22. Strobel,A., Issad,T., Camoin,L., Ozata,M. and Strosberg,A. (1998) A leptin missense mutation associated with hypogonadism and morbid obesity. *Nature Genetics*, **18**, 213–215.

23. Farooqi,I. and Matarese,G. (2002) Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *Journal of Clinical Investigation*, **110**, 1093–1103.

24. Farooqi,I. and Jebb,S. (1999) Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *New England Journal of Medicine*, **341**, 879–884.

25. Gibson,W.T., Farooqi,I.S., Moreau,M., DePaoli,A.M., Lawrence,E., O'Rahilly,S. and Trussell,R. a (2004) Congenital leptin deficiency due to homozygosity for the Delta133G mutation: report of another case and evaluation of response to four years of leptin therapy. *The Journal of Clinical Endocrinology and Metabolism*, **89**, 4821–4826.

26. Licinio,J., Caglayan,S., Ozata,M., Yildiz,B.O., de Miranda,P.B., O'Kirwan,F., Whitby,R., Liang,L., Cohen,P., Bhasin,S., et al. (2004) Phenotypic effects of leptin replacement on morbid obesity, diabetes mellitus, hypogonadism, and behavior in

leptin-deficient adults. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 4531–4536.

27. Clément,K., Vaisse,C., Lahlou,N., Cabrol,S., Pelloux,V., Cassuto,D., Gourmelen,M., Dina,C., Chambaz,J., Lacorte,J.M., et al. (1998) A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*, **392**, 398–401.

28. Farooqi,I.S., Wangensteen,T., Collins,S., Kimber,W., Matarese,G., Keogh,J.M., Lank,E., Bottomley,B., Lopez-Fernandez,J., Ferraz-Amaro,I., et al. (2007) Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *The New England Journal of Medicine*, **356**, 237–247.

29. Krude,H., Biebermann,H., Luck,W., Horn,R., Brabant,G. and Grüters,A. (1998) Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nature Genetics*, **19**, 155–157.

30. Krude,H. (2003) Obesity Due to Proopiomelanocortin Deficiency: Three New Cases and Treatment Trials with Thyroid Hormone and ACTH4-10. *Journal of Clinical Endocrinology & Metabolism*, **88**, 4633–4640.

31. Challis,B.G., Pritchard,L.E., Creemers,J.W.M., Delplanque,J., Keogh,J.M., Luan,J., Wareham,N.J., Yeo,G.S.H., Bhattacharyya,S., Froguel,P., et al. (2002) A missense mutation disrupting a dibasic prohormone processing site in pro-opiomelanocortin (POMC) increases susceptibility to early-onset obesity through a novel molecular mechanism. *Human Molecular Genetics*, **11**, 1997–2004.

32. Farooqi,I.S., Drop,S., Clements,A., Keogh,J.M., Biernacka,J., Lowenbein,S., Challis,B.G. and O'Rahilly,S. (2006) Heterozygosity for a POMC-null mutation and increased obesity risk in humans. *Diabetes*, **55**, 2549–2553.

33. Vaisse,C., Clement,K., Guy-grand,B. and Froguel,P. (1998) A frameshift mutation in human MC4R is associated with a dominant form of obesity WRN , is a 3 ′→ 5 ′ exonuclease. *Nature*, **20**, 113–114.

34. Yeo,G.S., Farooqi,I.S., Aminian,S., Halsall,D.J., Stanhope,R.G. and O'Rahilly,S. (1998) A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nature Genetics*, **20**, 111–112.

35. Hainerová,I., Larsen,L.H., Holst,B., Finková,M., Hainer,V., Lebl,J., Hansen,T. and Pedersen,O. (2007) Melanocortin 4 receptor mutations in obese Czech children: studies of prevalence, phenotype development, weight reduction response, and functional analysis. *The Journal of Clinical Endocrinology and Metabolism*, **92**, 3689–3696.

36. Farooqi,I.S., Keogh,J.M., Yeo,G.S.H., Lank,E.J., Cheetham,T. and O'Rahilly,S. (2003) Clinical Spectrum of Obesity and Mutations in the Melanocortin 4 Receptor Gene. *The New England Journal of Medicine*, **348**, 1085–1095.

37. Calton,M.A., Ersoy,B.A., Zhang,S., Kane,J.P., Malloy,M.J., Pullinger,C.R., Bromberg,Y., Pennacchio,L.A., Dent,R., McPherson,R., et al. (2009) Association of functionally significant Melanocortin-4 but not Melanocortin-3 receptor mutations with severe adult obesity in a large North American case-control study. *Human Molecular Genetics*, **18**, 1140–1147.

38. Lubrano-Berthelier,C. (2004) A homozygous null mutation delineates the role of the melanocortin-4 receptor in humans. *Journal of Clinical Endocrinology & Metabolism*, **89**, 2028–2032.

39. Dubern,B., Bisbis,S., Talbaoui,H., Le Beyec,J., Tounian,P., Lacorte,J.-M. and Clément,K. (2007) Homozygous null mutation of the melanocortin-4 receptor and severe early-onset obesity. *The Journal of Pediatrics*, **150**, 613–7, 617.e1.

40. Zakel,U., Wudy,S., Heinzel-Gutenbrunner,M., Gorg,T., Schafer,H., Gortner,L. and Blum,W. (2005) Prevalence of melanocortin 4 receptor (MC4R) mutations an polymorphisms consecutively ascertained obese children and adolescents from a pediatric health care utilization population. *Klin Padiatr*, **217**, 244–249.

41. Wang,C., Liang,L., Wang,H., Fu,J., Hebebrand,J. and Hinney,A. (2006) Several mutations in the melanocortin 4 receptor gene are associated with obesity in Chinese children and adolescents. *Journal of Endocrinological Investigation*, **29**, 894–898.

42. Holder,J., Butte,N. and Zinn,A. (2000) Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. *Human Molecular Genetics*, **9**, 101–108.

43. Hung,C.-C.C., Luan,J., Sims,M., Keogh,J.M., Hall,C., Wareham,N.J., O'Rahilly,S. and Farooqi,I.S. (2007) Studies of the SIM1 gene in relation to human obesity and obesity-related traits. *International Journal of Obesity*, **31**, 429–434.

44. Bonnefond,A., Raimondo,A., Stutzmann,F., Ghoussaini,M., Ramachandrappa,S., Bersten,D.C., Durand,E., Vatin,V., Balkau,B., Lantieri,O., et al. (2013) Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi–like features. *The Journal of Clinical Investigation*, **123**, 1–5.

45. Ramachandrappa,S., Raimondo,A., Cali,A.M.G., Keogh,J.M., Henning,E., Saeed,S., Thompson,A., Garg,S., Bochukova,E.G., Brage,S., et al. (2013) Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *Journal of Clinical Investigation*, **1**, 1–9.

46. Ahituv,N., Kavaslar,N., Schackwitz,W., Ustaszewska,A., Martin,J., Hebert,S., Doelle,H., Ersoy,B., Kryukov,G., Schmidt,S., et al. (2007) Medical sequencing at

the extremes of human body mass. *American Journal of Human Genetics*, **80**, 779–791.

47. Yamaki,A., Kudoh,J., Shimizu,N. and Shimizu,Y. (2004) A novel nuclear localization signal in the human single-minded proteins SIM1 and SIM2. *Biochemical and Biophysical Research Communications*, **313**, 482–488.

48. Chaudhary,J. and Skinner,M.K. (1999) Basic helix-loop-helix proteins can act at the E-box within the serum response element of the c-fos promoter to influence hormone-induced promoter activation in Sertoli cells. *Molecular endocrinology (Baltimore, Md.)*, **13**, 774–786.

49. Erbel,P.J. a, Card,P.B., Karakuzu,O., Bruick,R.K. and Gardner,K.H. (2003) Structural basis for PAS domain heterodimerization in the basic helix--loop--helix-PAS transcription factor hypoxia-inducible factor. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 15504–15509.

50. Nambu,J.F., Franks,R.G., Hu,S. and Crews,S.T. (1990) The single-minded Gene of Drosophila for the Expression of Genes Important Development of CNS Midline Cells Is Required for the. *Cell*, **63**, 63–75.

51. Zhou,L., Xiao,H. and Nambu,J.R. (1997) CNS midline to mesoderm signaling in Drosophila. *Mechanisms of Development*, **67**, 59–68.

52. House,S.B., Thomas,A., Kusano,K. and Gainer,H. (1998) Stationary organotypic cultures of oxytocin and vasopressin magnocellular neurones from rat and mouse hypothalamus. *Journal of Neuroendocrinology*, **10**, 849–861.

53. Serluca,F.C. and Fishman,M.C. (2001) Pre-pattern in the pronephric kidney field of zebrafish. *Development (Cambridge, England)*, **128**, 2233–2241.

54. Eaton,J.L. and Glasgow,E. (2006) The zebrafish bHLH PAS transcriptional regulator, single-minded 1 (sim1), is required for isotocin cell development. *Developmental Dynamics*, **235**, 2071–2082.

55. Michaud,J.L., Boucher,F., Melnyk,A., Gauthier,F., Goshu,E., Lévy,E., Mitchell,G. a, Himms-Hagen,J. and Fan,C.M. (2001) Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus. *Human Molecular Genetics*, **10**, 1465–1473.

56. Holder,J.L., Zhang,L., Kublaoui,B.M., DiLeone,R.J., Oz,O.K., Bair,C.H., Lee,Y.-H. and Zinn,A.R. (2004) Sim1 gene dosage modulates the homeostatic feeding response to increased dietary fat in mice. *American Journal of Physiology Endocrinology and Metabolism*, **287**, E105–E113.

57. Michaud,J.L., Rosenquist,T., May,N.R. and Fan,C.-M. (1998) Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1. *Genes & Development*, **12**, 3264–3275.

58. Michaud,J.L., DeRossi,C., May,N.R., Holdener,B.C. and Fan,C.M. (2000) ARNT2 acts as the dimerization partner of SIM1 for the development of the hypothalamus. *Mechanisms of Development*, **90**, 253–261.

59. Butler, a a, Marks,D.L., Fan,W., Kuhn,C.M., Bartolome,M. and Cone,R.D. (2001) Melanocortin-4 receptor is required for acute homeostatic responses to increased dietary fat. *Nature Neuroscience*, **4**, 605–611.

60. Balthasar,N., Dalgaard,L.T., Lee,C.E., Yu,J., Funahashi,H., Williams,T., Ferreira,M., Tang,V., McGovern,R. a, Kenny,C.D., et al. (2005) Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell*, **123**, 493–505.

61. Kublaoui,B.M., Holder,J.L., Gemelli,T. and Zinn,A.R. (2006) Sim1 haploinsufficiency impairs melanocortin-mediated anorexia and activation of paraventricular nucleus neurons. *Molecular Endocrinology*, **20**, 2483–2492.

62. Tolson,K.P., Gemelli,T., Gautron,L., Elmquist,J.K., Zinn,A.R. and Kublaoui,B.M. (2010) Postnatal Sim1 deficiency causes hyperphagic obesity and reduced Mc4r and oxytocin expression. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **30**, 3803–3812.

63. Yang,C., Gagnon,D., Vachon,P., Tremblay,A., Levy,E., Massie,B. and Michaud,J.L. (2006) Adenoviral-mediated modulation of Sim1 expression in the paraventricular nucleus affects food intake. *The Journal of Neuroscience*, **26**, 7116–7120.

64. Turleau,C., Demay,G., Cabanis,M.O., Lenoir,G. and de Grouchy,J. (1988) 6Q1 Monosomy: a Distinctive Syndrome. *Clinical Genetics*, **34**, 38–42.

65. Villa,A., Urioste,M., Bofarull,J.M. and Martinez-Frias,M.-L. (1995) De Novo Interstitial Deletion q16.2q21 on Chromsome 6. *American Journal of Human Geneticsjournal of medical genetics*, **383**, 379–383.

66. Gilhuis,H.J., Ravenswaaij,C.M.A.V.A.N., Hamel,B.E.N.J.C. and Ls,È. (2000) Interstitial 6q deletion with a Prader ± Willi-like phenotype : a new case and review of the literature Cytogenetic studies. *European Journal of Paediatric Neurology*, **4**, 39–43.

67. Faivre,L., Cormier-Daire,V., Lapierre,J., Colleaux,L., Jacquemont,S., Genevieve,D., Saunier,P., Munnich,A., Turleau,C., Romana,S., et al. (2002) Deletion of the SIM1 gene (6q16.2) in a patient with Prader-Willi-like phenotype. *Journal of Medical Genetics*, **39**, 594–596.

68. Varela,M.C., Simões-Sato,A.Y., Kim,C. a, Bertola,D.R., De Castro,C.I.E. and Koiffmann,C.P. (2006) A new case of interstitial 6q16.2 deletion in a patient with Prader-Willi-like phenotype and investigation of SIM1 gene deletion in 87 patients with syndromic obesity. *European Journal of Medical Genetics*, **49**, 298–305.

69. Bonaglia,M.C., Ciccone,R., Gimelli,G., Gimelli,S., Marelli,S., Verheij,J., Giorda,R., Grasso,R., Borgatti,R., Pagone,F., et al. (2008) Detailed phenotype-genotype study in five patients with chromosome 6q16 deletion: narrowing the critical region for Prader-Willi-like phenotype. *European Journal of Human Genetics*, **16**, 1443–1449.

70. Rosenfeld,J. a, Amrom,D., Andermann,E., Andermann,F., Veilleux,M., Curry,C., Fisher,J., Deputy,S., Aylsworth,A.S., Powell,C.M., et al. (2012) Genotype-phenotype correlation in interstitial 6q deletions: a report of 12 new cases. *Neurogenetics*, **13**, 31–47.

71. Ghoussaini,M., Stutzmann,F. and Couturier,C. (2010) Analysis of the SIM1 contribution to polygenic obesity in the French population. *Obesity*, **18**, 1670–1675.

72. Traurig,M., Mack,J., Hanson,R.L., Ghoussaini,M., Meyre,D., Knowler,W.C., Kobes,S., Froguel,P., Bogardus,C. and Baier,L.J. (2009) Common Variation in SIM1 Is Reproducibly Associated With BMI in Pima Indians. *Diabetes*, **58**, 1682–1689.

73. Visel,A., Akiyama,J., Shoukry,M. and Afzal,V. (2009) Functional autonomy of distant-acting human enhancers. *Genomics*, **93**, 509–513.

74. Woolfe,A., Goodson,M., Goode,D.K., Snell,P., McEwen,G.K., Vavouri,T., Smith,S.F., North,P., Callaway,H., Kelly,K., et al. (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biology*, **3**, e7.

75. Pennacchio,L. a, Ahituv,N., Moses,A.M., Prabhakar,S., Nobrega,M. a, Shoukry,M., Minovitsky,S., Dubchak,I., Holt,A., Lewis,K.D., et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, **444**, 499–502.

76. Visel,A., Prabhakar,S., Akiyama,J. a, Shoukry,M., Lewis,K.D., Holt,A., Plajzer-Frick,I., Afzal,V., Rubin,E.M. and Pennacchio,L. a (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature Genetics*, **40**, 158–160.

77. Banerji,J., Rusconi,S. and Schaffner,W. (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, **27**, 299–308.

78. Neznanov,N. (1997) A Regulatory Element within a Coding Exon Modulates Keratin 18 Gene Expression in Transgenic Mice. *Journal of Biological Chemistry*, **272**, 27549–27557.

79. Tümpel,S., Cambronero,F., Sims,C., Krumlauf,R. and Wiedemann,L.M. (2008) A regulatory module embedded in the coding region of Hoxa2 controls expression in rhombomere 2. *Proceedings of the ...*, **105**, 20077–20082.

80. Lettice,L. a. (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, **12**, 1725–1735.

81. Sagai,T., Hosoya,M., Mizushina,Y., Tamura,M. and Shiroishi,T. (2005) Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development (Cambridge, England)*, **132**, 797–803.

82. Sagai,T., Amano,T., Tamura,M., Mizushina,Y., Sumiyama,K. and Shiroishi,T. (2009) A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. *Development (Cambridge, England)*, **136**, 1665–1674.

83. VanderMeer,J.E. and Ahituv,N. (2011) Cis-Regulatory Mutations Are a Genetic Cause of Human Limb Malformations. *Developmental Dynamics*, **240**, 920–930.

84. Lomvardas,S., Barnea,G., Pisapia,D.J., Mendelsohn,M., Kirkland,J. and Axel,R. (2006) Interchromosomal interactions and olfactory receptor choice. *Cell*, **126**, 403–413.

85. Souza,F. de (2005) Identification of neuronal enhancers of the proopiomelanocortin gene by transgenic mouse analysis and phylogenetic footprinting. *Molecular and Cellular Biology*, **25**, 3076–3086.

86. Langlais,D., Couture,C., Sylvain-Drolet,G. and Drouin,J. (2011) A pituitary-specific enhancer of the POMC gene with preferential activity in corticotrope cells. *Molecular Endocrinology*, **25**, 348–359.

87. Bejerano,G., Lowe,C.B., Ahituv,N., King,B., Siepel,A., Salama,S.R., Rubin,E.M., Kent,W.J. and Haussler,D. (2006) A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, **441**, 87–90.

88. Birney,E., Stamatoyannopoulos,J. and Dutta,A. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.

89. Thomas,J. and Touchman,J. (2003) Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, **424**, 788–793.

90. Boffelli,D., Nobrega,M. a and Rubin,E.M. (2004) Comparative genomics at the vertebrate extremes. *Nature Reviews Genetics*, **5**, 456–465.

91. Dermitzakis,E., Reymond,A. and Antonarakis,S. (2005) Conserved non-genic sequences — an unexpected feature of mammalian genomes. *Nature Reviews Genetics*, **6**, 151–157.

92. Visel,A., Taher,L., Girgis,H., May,D., Golonzhka,O., Hoch,R. V, McKinsey,G.L., Pattabiraman,K., Silberberg,S.N., Blow,M.J., et al. (2013) A high-resolution enhancer atlas of the developing telencephalon. *Cell*, **152**, 895–908.

93. Kim,M.J., Skewes-Cox,P., Fukushima,H., Hesselson,S., Yee,S.W., Ramsey,L.B., Nguyen,L., Eshragh,J.L., Castro,R. a, Wen,C.C., et al. (2011) Functional characterization of liver enhancers that regulate drug-associated transporters. *Clinical Pharmacology and Therapeutics*, **89**, 571–578.

94. Arnosti,D.N. and Kulkarni,M.M. (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry*, **94**, 890–898.

95. Sandelin,A., Bailey,P., Bruce,S., Engström,P.G., Klos,J.M., Wasserman,W.W., Ericson,J. and Lenhard,B. (2004) Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC genomics*, **5**, 99.

96. Peterson,B.K., Hare,E.E., Iyer,V.N., Storage,S., Conner,L., Papaj,D.R., Kurashima,R., Jang,E. and Eisen,M.B. (2009) Big genomes facilitate the comparative identification of regulatory elements. *PLoS one*, **4**, e4688.

97. Ahituv,N., Prabhakar,S., Poulin,F., Rubin,E.M. and Couronne,O. (2005) Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. *Human Molecular Genetics*, **14**, 3057–3063.

98. Ghoussaini,M., Vatin,V., Lecoeur,C., Abkevich,V., Younus,A., Samson,C., Wachter,C., Heude,B., Tauber,M., Tounian,P., et al. (2007) Genetic study of the melanin-concentrating hormone receptor 2 in childhood and adulthood severe obesity. *The Journal of Clinical Endocrinology and Metabolism*, **92**, 4403–4409.

99. Ovcharenko,I., Nobrega,M.A., Loots,G.G. and Stubbs,L. (2004) ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic acids research*, **32**, W280–W286.

100. Kent,W., Sugnet,C., Furey,T., Roskin,K., Pringle,T., Zahler,A. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Research*, **12**, 996–1006.

101. Li,Q., Ritter,D., Yang,N., Dong,Z., Li,H., Chuang,J. and Guo,S. (2009) A systematic approach to identify functional motifs within vertebrate developmental enhancers. *Developmental Biology*, **337**, 484–495.

102. Fisher,S., Grice,E., Vinton,R., Bessling,S., Urasaki,A., Kawakami,K. and AS,M. (2006) Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nature Protocols*, **1**, 1297–1305.

103. White,R.M., Sessa,A., Burke,C., Bowman,T., LeBlanc,J., Ceol,C., Bourque,C., Dovey,M., Goessling,W., Burns,C.E., et al. (2008) Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell*, **2**, 183–189.

104. Eaton,J.L. and Glasgow,E. (2007) Zebrafish orthopedia (otp) is required for isotocin cell development. *Development Genes and Evolution*, **217**, 149–158.

105. Kothary,R., Clapoff,S., Brown,A. and Campbell,R. (1988) A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature*, **335**, 435–437.

106. Nagy,A., Gertsenstein,M., Vintersten,K. and Behringer,R. (2003) Manipulating the mouse embryo: a laboratory manual 3rd ed. Cold Spring Harbor Press, Cold Spring Harbor, NY.

107. Fan,C., Kuwana,E., Bulfone,A., Fletcher,C.F., Copeland,N.G., Jenkins,N.A., Crews,S., Martinez,S., Puelles,Ø.L., Rubenstein,J.L.R., et al. (1996) Expression Patterns of Two Murine Homologs of. *Molecular and Cellular Neuroscience*, **16**, 1–16.

108. Marion,J.-F., Yang,C., Caqueret,A., Boucher,F. and Michaud,J.L. (2005) Sim1 and Sim2 are required for the correct targeting of mammillary body axons. *Development*, **132**, 5527–5537.

109. Shimogori,T., Lee,D. a, Miranda-Angulo,A., Yang,Y., Wang,H., Jiang,L., Yoshida,A.C., Kataoka,A., Mashiko,H., Avetisyan,M., et al. (2010) A genomic atlas of mouse hypothalamic development. *Nature neuroscience*, **13**, 767–775.

110. Visel,A., Minovitsky,S., Dubchak,I. and Pennacchio,L. a (2007) VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Research*, **35**, D88–D92.

111. GENSAT *The Gene Expression Nervous System Atlas (GENSAT) Project, NINDS Contracts N01NS02331 & HHSN271200723701C to the Rockefeller University (New York, NY).*

112. Visel,A., Blow,M., Li,Z., Zhang,T. and Akiyama,J. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**, 854–858.

113. Abecasis,G.R., Auton,A., Brooks,L.D., DePristo,M. a, Durbin,R.M., Handsaker,R.E., Kang,H.M., Marth,G.T. and McVean,G. a (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.

114. Overstreet,L.S., Hentges,S.T., Bumaschny,V.F., de Souza,F.S.J., Smart,J.L., Santangelo,A.M., Low,M.J., Westbrook,G.L. and Rubinstein,M. (2004) A transgenic marker for newly born granule cells in dentate gyrus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **24**, 3251–3259.

115. Hong,J.-W., Hendrix,D. a and Levine,M.S. (2008) Shadow enhancers as a source of evolutionary novelty. *Science*, **321**, 1314.

116. Kawakami,K. (2005) Transposon tools and methods in zebrafish. *Developmental Dynamics*, **234**, 244–254.

117. Furey,T.S. (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nature Reviews Genetics*, **13**, 840–852.

118. Haberle,V. and Lenhard,B. (2012) Dissecting genomic regulatory elements in vivo. *Nature Biotechnology*, **30**, 504–606.

119. Westerfield,M. (2007) The Zebrafish Book, A guide for the laboratory use of zebrafish (Danio rerio) 5th ed. University of Oregon, Press, Eugene.

120. Nüsslein-Volhard, C. and Dahm,R. (2002) Zebrafish, A Practical Approach Oxford University Press.

121. Schmidt,A., Tief,K., Foletti,A., Hunziker,A., Penna,D., Hummler,E. and Beermann,F. (1998) lacZ transgenic mice to monitor gene expression in embryo and adult. *Brain Research Protocols*, **3**, 54–60.

122. Flandin,P., Zhao,Y., Vogt,D., Jeong,J. and Long,J. (2011) Lhx6 and Lhx8 Coordinately Induce Neuronal Expression of Shh that Controls the Generation of Interneuron Progenitors. *Neuron*, **70**, 939–950.

123. Conlon,R. and Rossant,J. (1992) Exogenous retinoic acid rapidly induces anterior ectopic expression of murine Hox-2 genes in vivo. *Development*, **116**, 357–368.

124. Matys,V., Kel-Margoulis,O. V, Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K., et al. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research*, **34**, D108–D110.

125. Quandt,K., Frech,K., Karas,H., Wingender,E. and Werner,T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Research*, **23**, 4878–84.

126. Loots,G.G., Ovcharenko,I., Pachter,L., Dubchak,I. and Rubin,E.M. (2002) rVista for Comparative Sequence-Based Discovery of Functional Transcription Factor Binding Sites. *Genome Research*, **12**, 832–839.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

8/25/13

Date