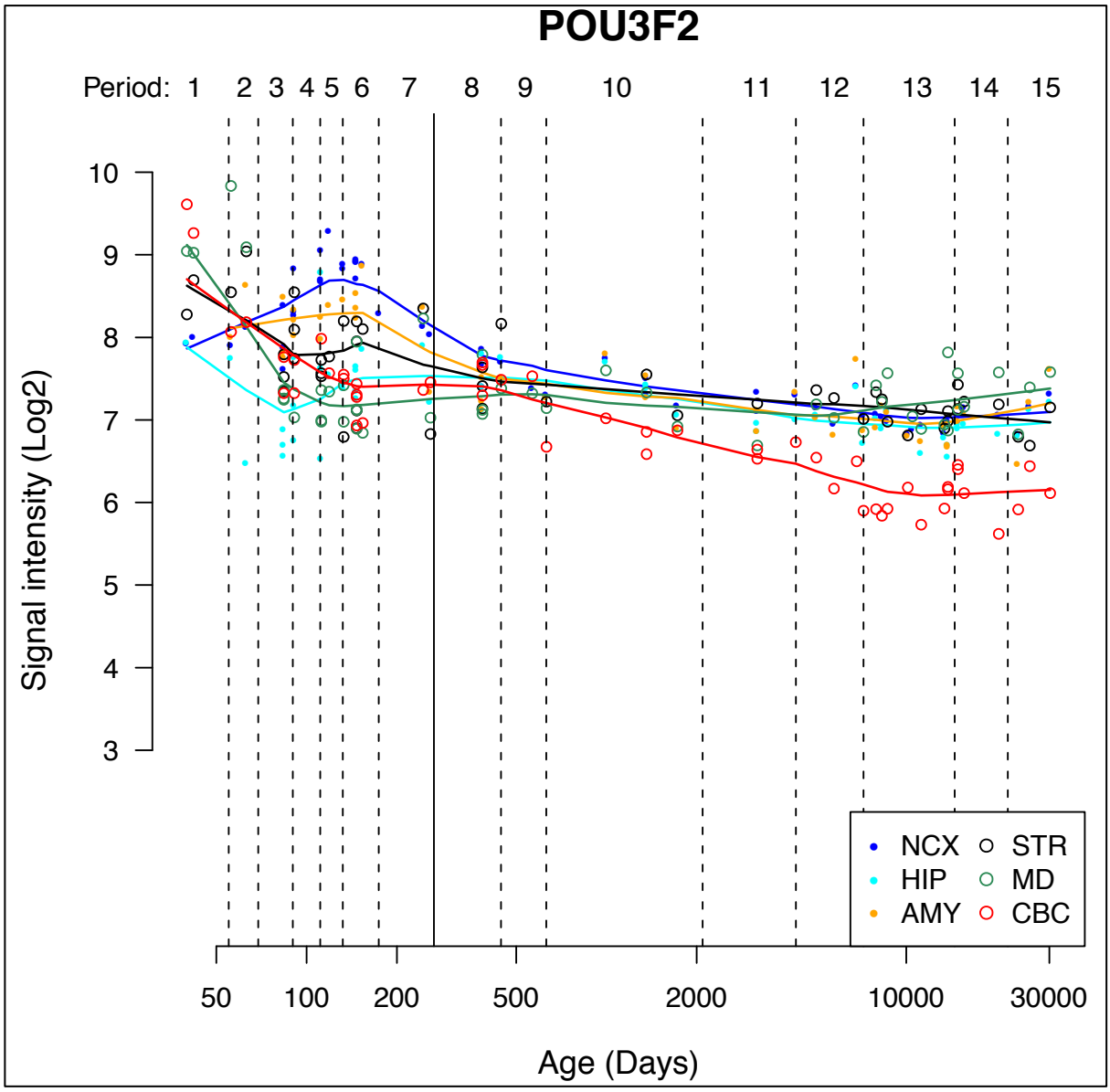
The American Journal of Human Genetics

Supplemental Data

Small 6q16.1 Deletions Encompassing *POU3F2* Cause Susceptibility to Obesity and Variable Developmental Delay with Intellectual Disability

Paul R. Kasher, Katherine E. Schertz, Megan Thomas, Adam Jackson, Silvia Annunziata, María J. Ballesta-Martinez, Philippe M. Campeau, Peter E. Clayton, Jennifer L. Eaton, Tiziana Granata, Encarna Guillén-Navarro, Cristina Hernando, Caroline E. Laverriere, Agne Liedén, Olaya Villa-Marcos, Meriel McEntagart, Ann Nordgren, Chiara Pantaleoni, Céline Pebrel-Richard, Catherine Sarret, Francesca L. Sciacca, Ronnie Wright, Bronwyn Kerr, Eric Glasgow, and Siddharth Banka

Figure S1. *POU3F2* is highly expressed throughout fetal and adult human brain.
Figure generated from Human Brain Transcriptome project data (http://hbatlas.org/) [1] showing high spatio-temporal expression of *POU3F2*.
The legends and axes are labeled and are self-explanatory.



Key: *NCX neocortex; HIP hippocampus; AMY amygdala; STR striatum; MD Mediodorsal nucleus of the thalamus; CBC cerebellar cortex.*

Figure S2: High level of expression of *POU3F2* and related genes in the human hypothalamus.

These figures were generated using Brain Explorer tool based on data from the Allen Human Brain Atlas (http://human.brain-map.org/) [2] in following steps –

(i) Brain Explorer tool was downloaded from http://human.brain-map.org/static/brainexplorer and installed on a local computer.

(ii) Allen Human Brain Atlas (37MB) – October 2013 was installed from http://www.brain-map.org

(iii) Gene name was entered in the search box at URL - http://human.brain-map.org/microarray/search (with settings: Resolution = structures; Color Map = z-score)

(iv) Relevant heat map was selected by clicking that activates the Brain Explorer link on the webpage.

(v) Clicking on the activated link Brain Explorer link opens the Application and enables visualization of spatial rendition of gene expression.

(vi) Within the application specific brain structures of interest were selected to be highlighted and to restrict the visualized expression points (in this case the hypothalamus and the hippocampal formation).

(vii) Various brain sections were visualized and a representative image was selected and exported as an image file.

Figures are arranged to represent the Leptin>MC4R>SIM1>POU3F2>OXT pathway and show that all five genes are highly expressed in the human hypothalamus. The bottom left figure, represents expression pattern of *POU3F3*. Comparing expression patterns of human *POU3F3* and *POU3F2* shows that it is similar to what has been observed in mice [3].

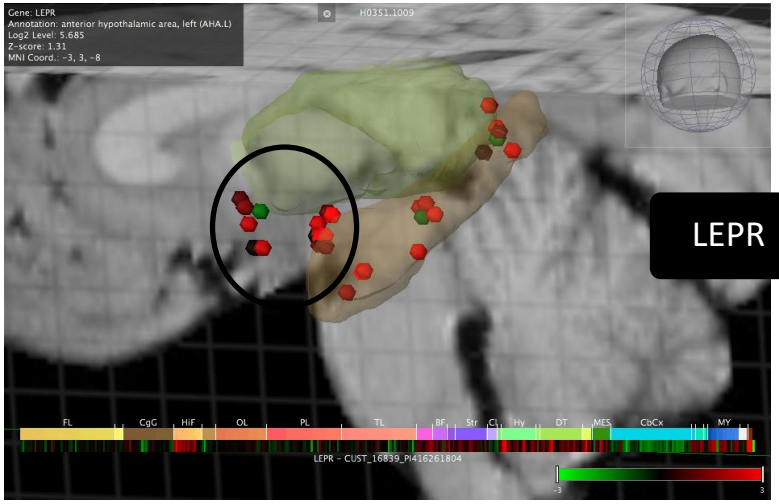
In all the figures the inset at the top left corner provides the gene name, anatomical centre of the figure, Log2 level, Z-score and co-ordinates.

The top right corner inset shows a 'compass' to aid orientation of the main figure.

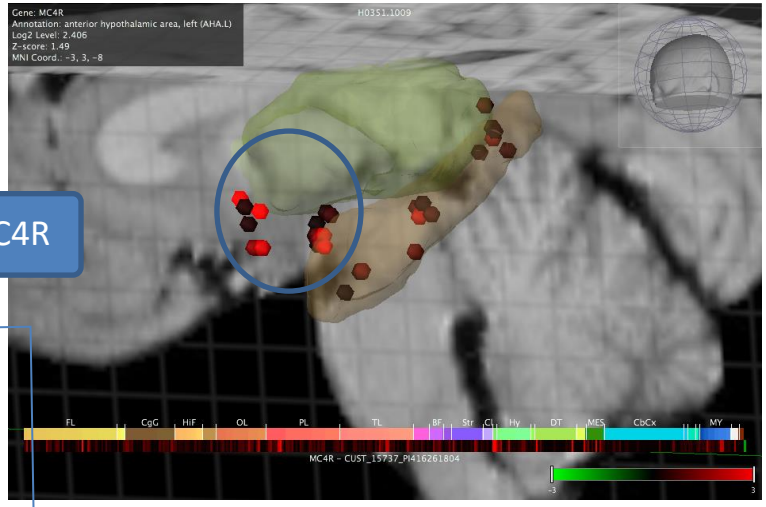
The hypothalamic area is circled, transparent green area represents the thalamus and transparent orange area represents the hippocampal formation.

The expression level at each measured point is given on a green-red scale (bottom right). The two horizontal bars at the bottom provide more detailed information on the expression of the particular gene via the green-red heat map in each of the following regions of the brain – FL (frontal lobe), CgC (cingulate gyrus), HiF (hippocampal formation), OL (occipital lobe), PL (parietal lobe), TL (temporal lobe), BF (basal forebrain), Str (striatum), C (claustrum), Hy (hypothalamus), DT (dentate nucleus), MES (mesencephalon), CbCx (cerebellar cortex) and My (myelencephalon).

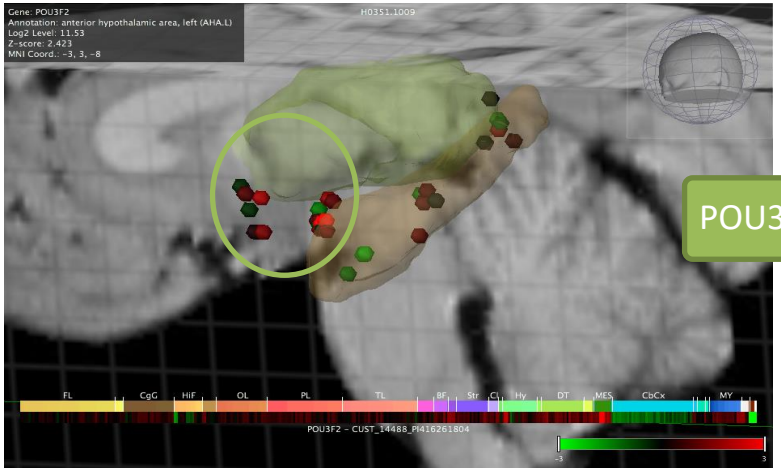
The anatomical section and the donor sample identifiers are provided at the top centre and bottom centre respectively.



LEPR

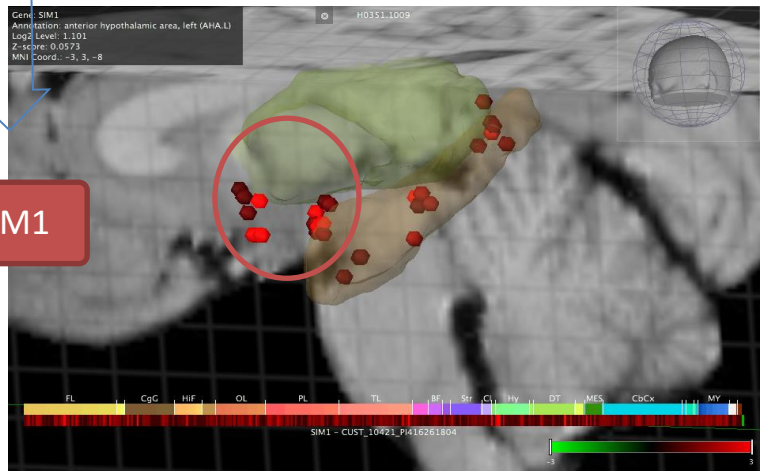


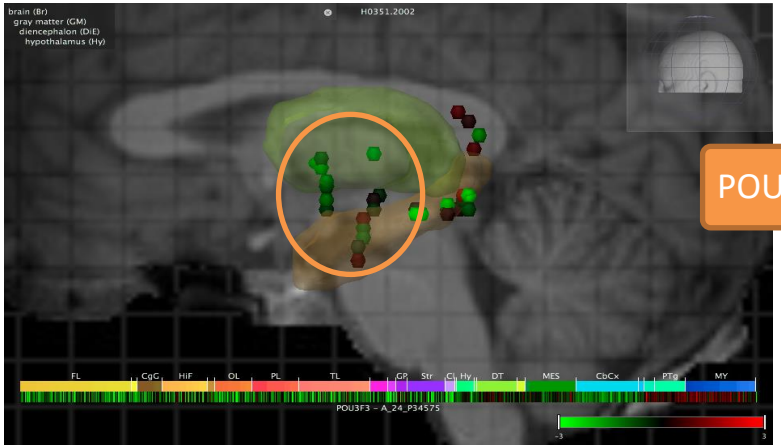
MC4R



POU3F2

SIM1





POU3F3

OXT

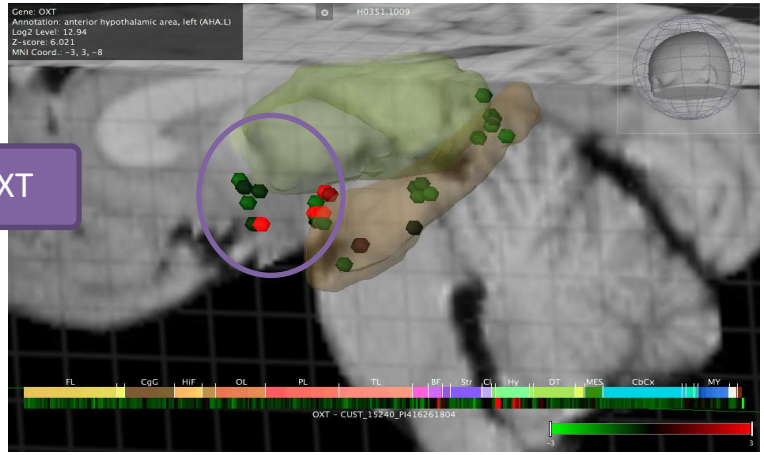
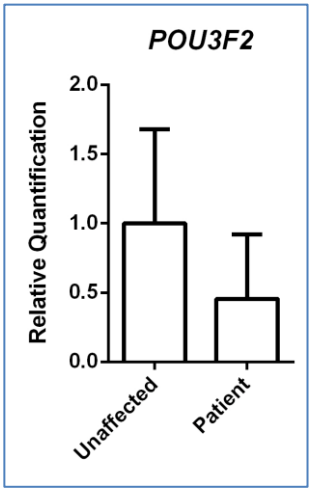


Figure S3: *POU3F2* quantitative reverse transcription PCR in patient cell lines.

POU3F2 expression was assessed in Epstein-Barr virus transformed lymphoblastoid cell lines (LCLs) derived from a patient and unaffected sibling (Family 1). LCLs were maintained as described previously[4]. Total RNA was extracted from LCLs ($10x10^6$ cells per sample) using TRIzol reagent (Thermo Fisher). RNA concentration was assessed using a spectrophotometer (FLUOstar Omega, Labtech). cDNA was synthesised from 1600ng RNA using the High Capacity RNA to cDNA kit (Life Technologies, UK). Quantitative reverse transcription PCR (qRT-PCR) analysis was performed using the TaqMan Universal PCR Master Mix (Applied Biosystems). Each PCR was performed three times in triplicate per cDNA sample (total of 9 technical repeats per sample). The relative abundance of *POU3F2* transcript was measured using a *POU3F2* Taqman probe (Hs00271595_s1) following normalisation to the expression level of two 'housekeeper' genes, *HPRT1* (Hs03929096_g1) and *18s* (Hs999999001_s1), and assessed with the Applied Biosystems StepOne Software v2.1. Statistical significance between groups was determined by t-tests using DataAssist v2.0 (Applied Biosystems). Fold change values were plotted using Graphpad Prism v5.0 and defined as the Relative Quantification (RQ).



An approximate 50% reduction in POU3F2 expression in LCLs derived from patient in comparison to his unaffected sibling (p=0.1447).
Amplification of the POU3F2 transcript occurred at very late cycle numbers, suggesting that this gene is expressed at very low levels in peripheral lymphocytes.
Indeed, we hypothesise that low expression is the most likely cause for the high variation observed, even after 9 technical repeats and using a high concentration of cDNA template.

Figure S4: Human and zebrafish POU3F2 genes are highly conserved.

Protein sequence alignment was created using Colbalt Constraint-based Multiple Protein Alignment Tool. The sequences correspond to zebrafish Pou3f2a (NP_571364), zebrafish Pou3f2b (NP_571235.1), human POU3F2 (accession number: NP_005595), and mouse POU3F2* (accession number: NP_032925.1), human POU3F1 (NP_002690), human POU3F3 (NP_006227), and human POU3F4 (NP_000298).

The entire protein sequence is highly conserved among mice, humans, and both orthologs in zebrafish. The POU homeodomain (amino acids 95-251; Pou3f2a sequence), highlighted in red, has 92% identity and 97% similarity for all seven sequences.

FIGURE 3

```
Pou3f2a1     ----LSH---   -----------   ------GgEGSPWPA   SPLGEQDIKPV-E---E------------LQ   H  28
Pou3f2b1     ---ALSH---   -----------   ------G-EGGPWSS   SPLGEQDIKPA-V---QSPR-DEMHNSSNLQ   H  39
POU3F2 1     witALSHGGG   GGGGGGGGGGG   GGGGGGG-DGSPWST   SPLGQPDIKPS-VvVQQGGRGDELHGPGALQ   Q  66
POU3F2*1     witALSHGGG   GGGGGGGGGGG   GGGGGGG-DGSPWST   SPLGQPDIKPS-VvVQQGGRGDELHGPGALQ   Q  66
POU3F1 1     ---GAGHPVG[9]GGGGGGDWAGG[5]GKAGGGG-TGR----   --------------ADDGGGGGGFHA-----   -  54
POU3F3 1     witALSHGGG   GGGGGGGGGGG   GGGGGGG-DGSPWST   SPLGQPDIKPS-VvVQQGGRGDELHGPGALQ   Q  66
POU3F4 1     ----------   -----------   -----LS-DGGPWSS[4]SPLDQQDVKPGrE---D------------LQ[7]R  37

Pou3f2a29    S    P--RQAHLVPQSQ--HHETAAWRATTTA-HMP-SMSTSNGqSL   IYSQP---   --------GY--G--EMH  77
Pou3f2b40    Q    S--RPPHLVHQTHGNHHDSRAWRTTTAA-HIP-SMATSNGqSL   IYSQPSFS   VNGLIPGSGQ--G--IHH  101
POU3F2 67    Q[21]Q--RPPHLVHHAANHHPGPGAWRSAAAAAHLPPSMGASNG-GL   LYSQPSFT   VNGMLGAGGQPAG--LHH  152
POU3F2*67    Q[21]QqqRPPHLVHHAANHHPGPGAWRSAAAAAHLPPSMGASNG-GL   LYSQPSFT   VNGMLGAGGQPAG--LHH  154
POU3F1 55    -    ------RLVHQGAAH--AGAAWAQGSTAHHLGPAMSPSPG-AS[9]LYAQAAYP[6]LAGMLAAGGGGAGpgLHH  129
POU3F3 67    Q[21]Q--RPPHLVHHAANHHPGPGAWRSAAAAAHLPPSMGASNG-GL   LYSQPSFT   VNGMLGAGGQPAG--LHH  152
POU3F4 38    S    P--HVAHHSPHTN--HPN--AWGASPAP-N-P-SI-TSSGqPL[1]VYSQP---   --------GF--T--VSG  83

Pou3f2a78    H------EEhhS    PHLSE---HG   HPQSLH---------------------------SDEDTPTSDDLEQFAKQF  114
Pou3f2b102   HSMRDAHEDhhS    PHLSD---HG   HPPSQHQHQSHQ-------------SHHD---HSDEDTPTSDDLEQFAKQF  155
POU3F2 153   HGLRDAHDE---    PHHAD---HH   PHPHSHPHQQPPPPPPPQGPPGHPGAHHD--PHSDEDTPTSDDLEQFAKQF  217
POU3F2*155   HGLRDAHDE---    PHHAD---HH   PHPHSHPHQQPPPPPPPQGPPGHPGAHHD--PHSDEDTPTSDDLEQFAKQF  219
POU3F1 130   ALHEDGHEA--Q[ 5]PPHLG---AH   GHAHGHAHAGGLHAAAAHLHPGAGGGGSSvgEHSDEDAPSSDDLEQFAKQF  202
POU3F3 153   HGLRDAHDE---    PHHAD---HH   PHPHSHPHQQPPPPPPPQGPPGHPGAHHD--PHSDEDTPTSDDLEQFAKQF  217
POU3F4 84    M------LEhgG[14]PVLREppdHG[4]HHCQDH---------------------------SDEETPTSDELEQFAKQF  141

Pou3f2a115   KQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLNKWLEEADSTSGSPTSLDKIAAQGRKRKK  194
Pou3f2b156   KQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLNKWLEEADSTSGSPTSLDKIAAQGRKRKK  235
POU3F2 218   KQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLNKWLEEADSSSGSPTSIDKIAAQGRKRKK  297
POU3F2*220   KQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLNKWLEEADSSSGSPTSIDKIAAQGRKRKK  299
POU3F1 203   KQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLNKWLEETDSSSGSPTNLDKIAAQGRKRKK  282
POU3F3 218   KQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLNKWLEEADSSSGSPTSIDKIAAQGRKRKK  297
POU3F4 142   KQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLNKWLEEADSSTGSPTSIDKIAAQGRKRKK  221

Pou3f2a195   RTSIEVSVKGALESHFLKCPKPGASEINSLADSLQLEKEVVRVWFCNRRQKEKRMTPQNGP-MAGSEDVY   G---DAS  267
Pou3f2b236   RTSIEVSVKGALESHFLKCPKPAASEITSLADSLQLEKEVVRVWFCNRRQKEKRMTPPGGP-LPGTEDVY   G---DTP  308
POU3F2 298   RTSIEVSVKGALESHFLKCPKPSAQEITSLADSLQLEKEVVRVWFCNRRQKEKRMTPPGGT-LPGAEDVY   GGSRDTP  373
POU3F2*300   RTSIEVSVKGALESHFLKCPKPSAQEITSLADSLQLEKEVVRVWFCNRRQKEKRMTPPGGT-LPGAEDVY   GGSRDTP  375
POU3F1 283   RTSIEVGVKGALESHFLKCPKPSAHEITGLADSLQLEKEVVRVWFCNRRQKEKRMTPAAGAgHPPMDDVY[7]GGGGASP  366
POU3F3 298   RTSIEVSVKGALESHFLKCPKPSAQEITSLADSLQLEKEVVRVWFCNRRQKEKRMTPPGGT-LPGAEDVY   GGSRDTP  373
POU3F4 222   RTSIEVSVKGVLETHFLKCPKPAAQEISSLADSLQLEKEVVRVWFCNRRQKEKRMTPP------------   G---DQQ  283

Pou3f2a268   P    H   HGAQTPVP     277
Pou3f2b309   P    H   HGVQTPVQ     318
POU3F2 374   P    H   HGVQTPVQ     383
POU3F2*376   P    H   HGVQTPVQ     385
POU3F1 367   P[11]H[4]HTLPGSVQ     391
POU3F3 374   P    H   HGVQTPVQ     383
POU3F4 284   P    H[4]HTVKTDTS[4]  301

*POU3F2 mouse
```

Figure S5: Human POU3F2 is orthologous with zebrafish Pou3f2a and Pou3f2b.
A phylogeny tree was created using MUSCLE Multiple Sequence Alignment. Protein sequences for all four human POU3 family proteins, POU3F1 (NP_002690), POU3F2 (NP_005595), POU3F3 (NP_006227), POU3F4 (NP_000298), as well as sequences for zebrafish Pou3f2a (accession number: NP_571364) and Pou3f2b (accession number: NP_571235.1), were used as input data. The output data cluster human POU3F2 more closely with zebrafish pou3f2a and pou3f2b than the other POU3 family proteins confirming that these genes are orthologous.

Human *POU3F2*, and zebrafish *pou3f2a,* and *pou3f2b* contain single protein coding exons, as well as conserved synteny (Synteny Database, http://syntenydb.uoregon.edu) (data not shown). The conserved synteny of human *POU3F2* with both *pou3f2a* and *pou3f2b* gene regions indicates that chromosomal duplication likely generated the zebrafish *pou3f2* paralogs.

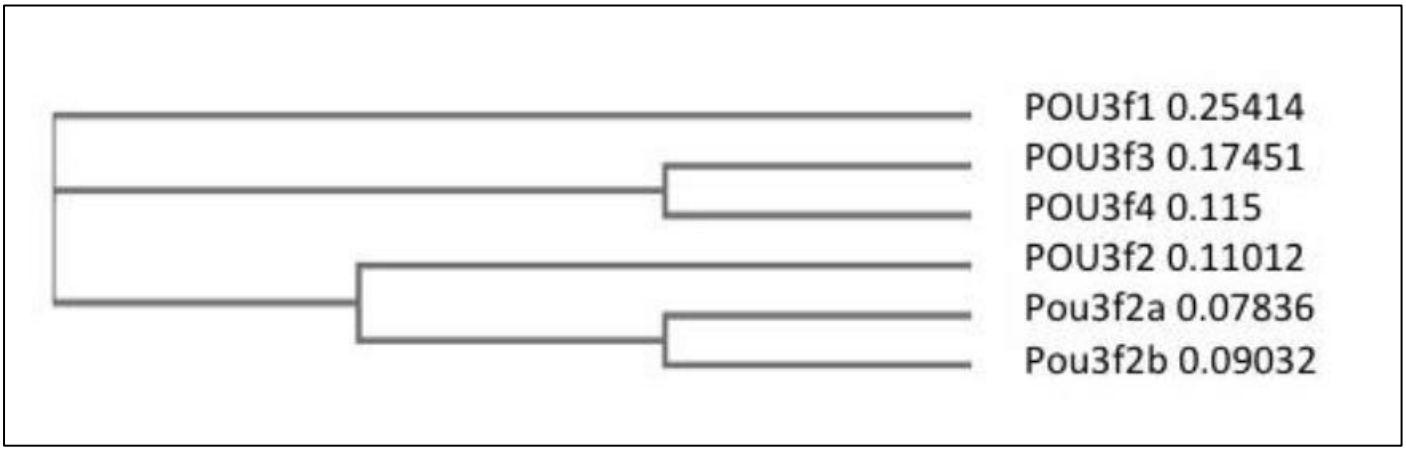
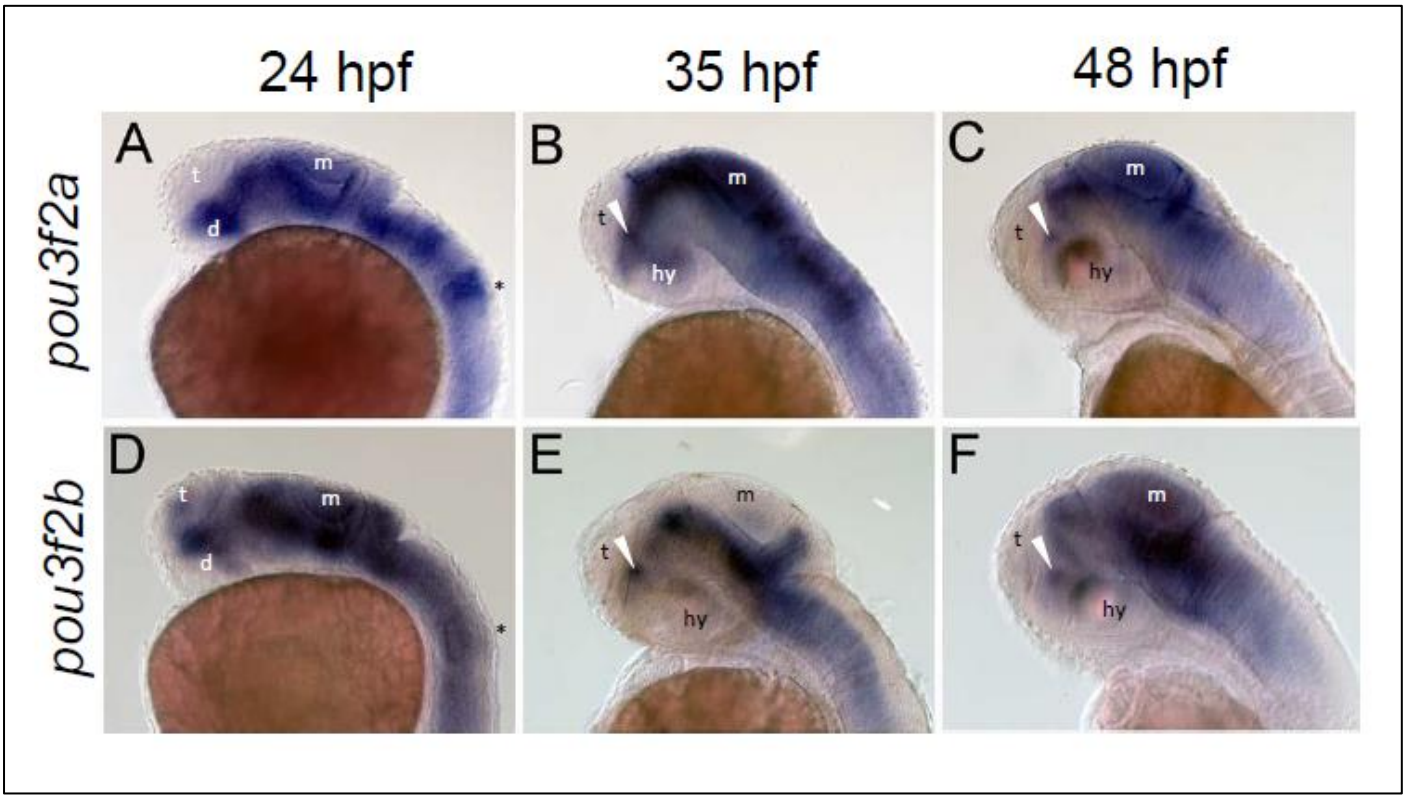


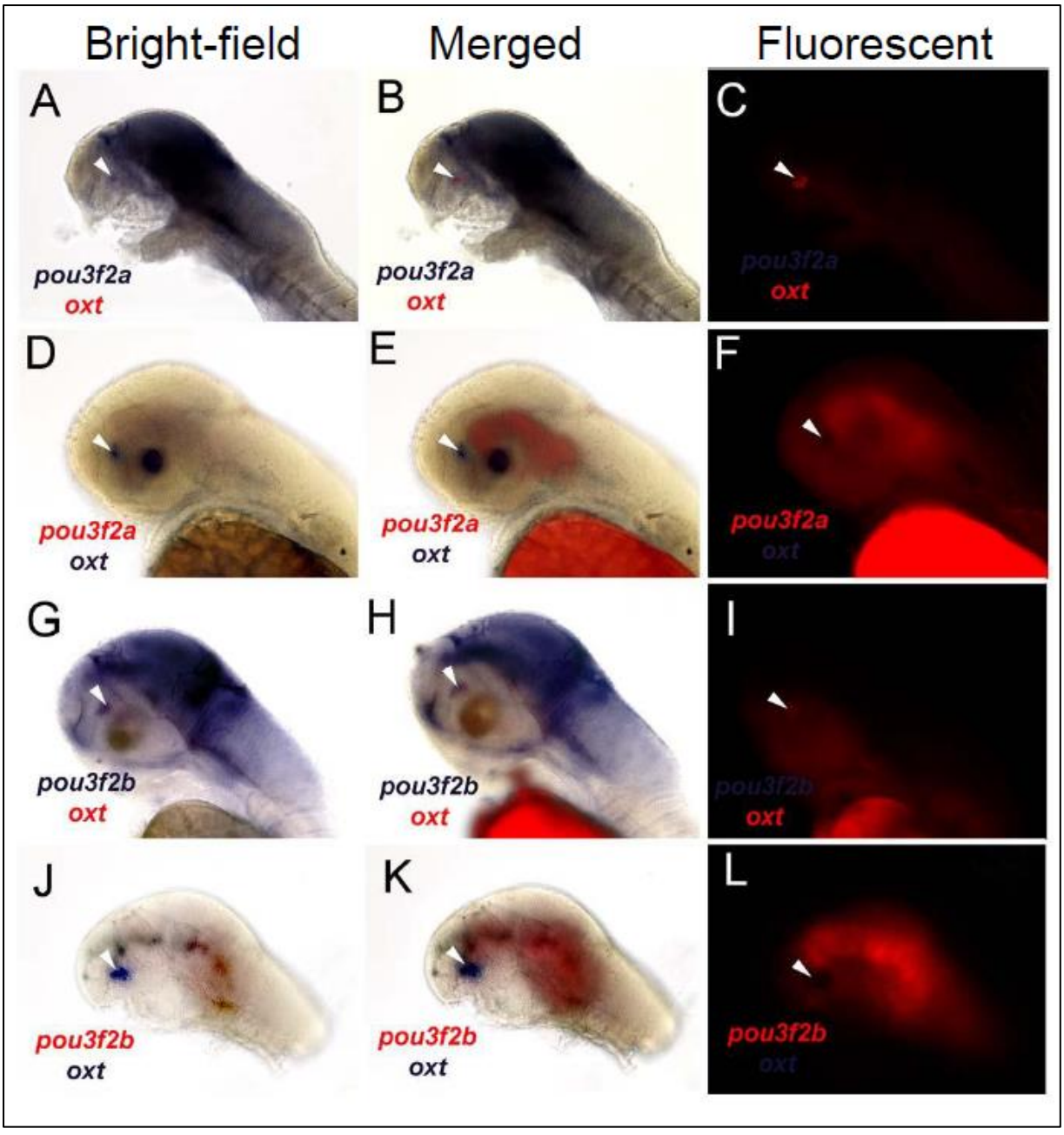
Figure S6: Developmental expression patterning of zebrafish *pou3f2a* and *pou3f2b.*
The expression patterns of *pou3f2a* and *pou3f2b* mRNAs were determined by Whole mount in situ hybridization (WISH) at 24, 35, and 48 hours post fertilization (hpf).
Lateral views, with anterior left, and dorsal up.



*Key: t, telencephalon; d, diencephalon; hy, hypothalamus; m, midbrain; *, marks the position of rhombomere r5.*

(**A, D**) At 24 hpf *pou3f2a* and *pou3f2b* are extensively expressed in the diencephalon, midbrain, and hindbrain. In the hindbrain, *pou3f2a* expression is concentrated in rhombomeres r1, r3 and r5.
(**B, E**) By 35 hpf, *pou3f2a* and *pou3f2b* expression becomes weaker in the anterior diencephalon along the telencephalic-diencephalic border, and in the posterior hypothalamus. Within the anterior diencephalic domain, strong *pou3f2a* expression is maintained in the neuroendocrine preoptic area (NPO), indicated with white arrowheads.
(**C, F**) By 48 hpf, *pou3f2a* and *pou3f2b* expression is seen in the diencephalon, the midbrain tegmentum and throughout the hindbrain. In the diencephalon, expression further narrows and strong expression of *pou3f2a* and *pou3f2b* remains restricted to a small area of the NPO.

Figure S7: Co-localization of *pou3f2a* and *pou3f2b* mRNAs with subsets of oxytocin cells.
Co-expression of *pou3f2a*, *pou3f2b* and *oxt* mRNA was determined by double label WISH in 48 hpf embryos. *Pou3f2a* and *pou3f2b* probes were labelled with digoxygenin (DIG) while oxytocin probes were labeled with fluorescein (FL). DIG and FL were detected using either BM Purple or Red Fast Stain.
The white arrowhead indicates the location of the *oxt* expression domain. (**A, D, G, J**) Bright field images (**C, F, I, L**) are the corresponding epifluorescence images, and (**B, E, H, K**) are the merged bright field and epifluorescence images.



(**A-C**) *pou3f2a* expression is visualized with purple staining and *oxt* expression is visualized with red staining. The eyes have been removed to better visualize the staining patterns. The purple *pou3f2a* staining quenches the red *oxt* stain, except in a small posterior region, of the *oxt* domain.
(**D-F**) The chromogenic staining is reversed such that *pou3f2a* expression is visualized with red staining and

oxt expression is visualized with purple staining. The purple *oxt* stain quenches the entire red *pou3f2a* stain in this region. Thus, *pou3f2a* appears to be co-expressed in a subset of oxytocin cells.
(**G-I**) *pou3f2b* expression is visualized with purple staining and *oxt* expression is visualized with red staining. The purple *pou3f2b* staining quenches the red *oxt* stain of the *oxt* domain, except for in a small posterior region as seen with *pou3f2a*.
(**J-L**) The chromogenic staining is reversed such that *pou3f2b* expression is visualized with red staining and *oxt* expression is visualized with purple staining. The eyes have been removed to better visualize the staining patterns. The purple *oxt* stain quenches the entire red *pou3f2b* stain in this region. Thus, similar to *pou3f2a*, *pou3f2b* appears to be co-expressed in a subset of oxytocin cells.

Figure S8: Knockdown of *pou3f2a* or *pou3f2b* has no effect on expression of *sim1a*

MO knockdown of *pou3f2a* or *pou3f2b* causes no obvious change in *sim1a* expression patterns. (**A-C**) Lateral views, 48 hpf embryos stained for *sim1a* expression by WISH. The location of the NPO is indicated with white arrowheads. (**A**) control MO injected embryos (**B**) *pou3f2a* MO injected embryo, (**C**) *pou3f2b* MO injected embryo.

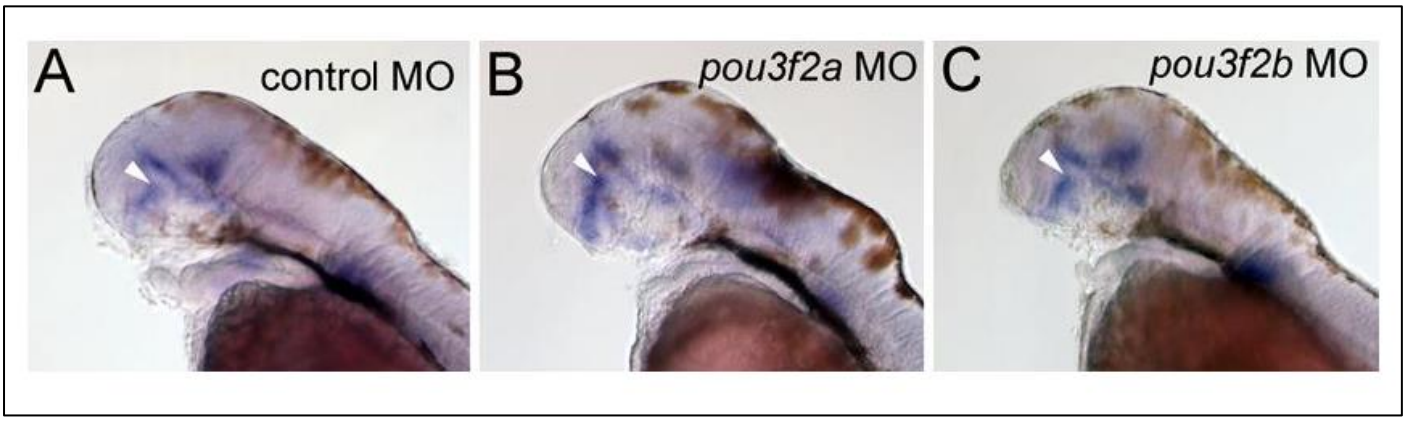
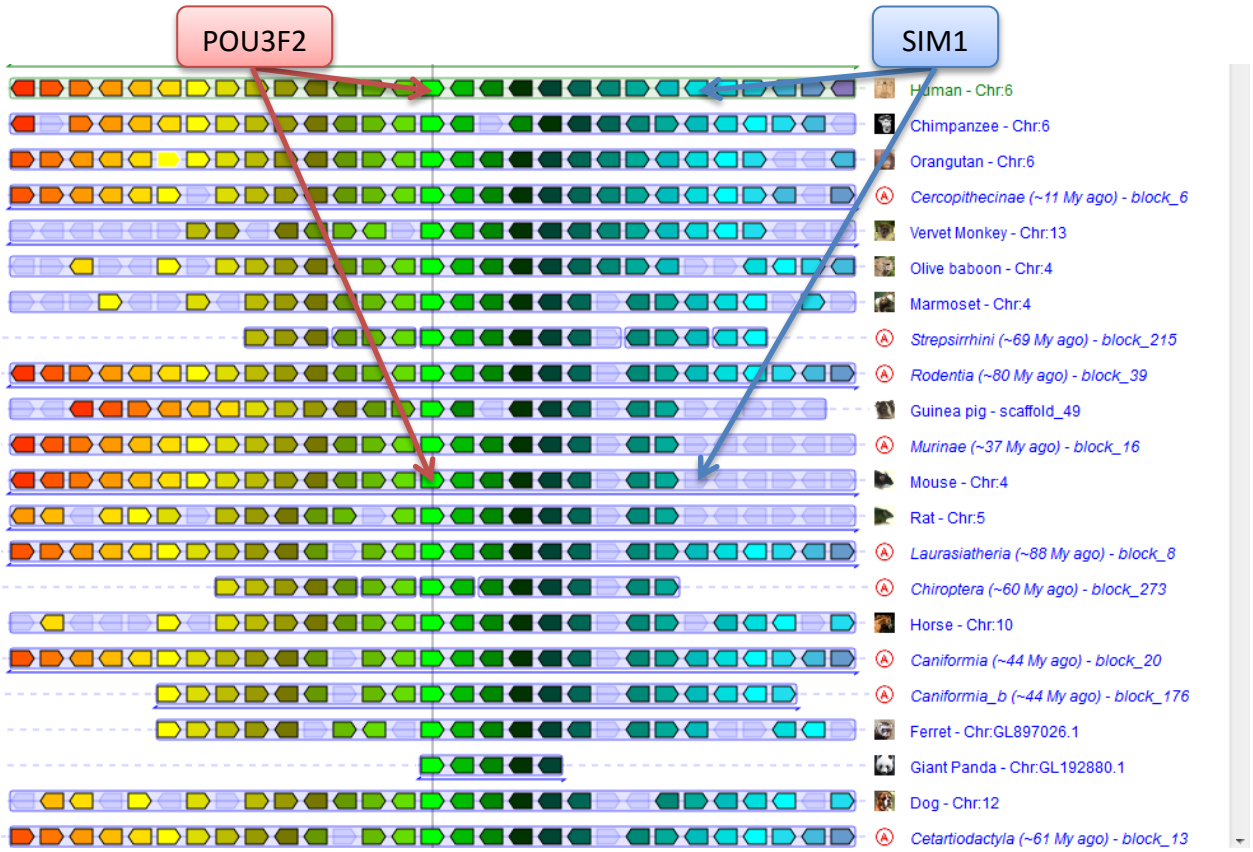


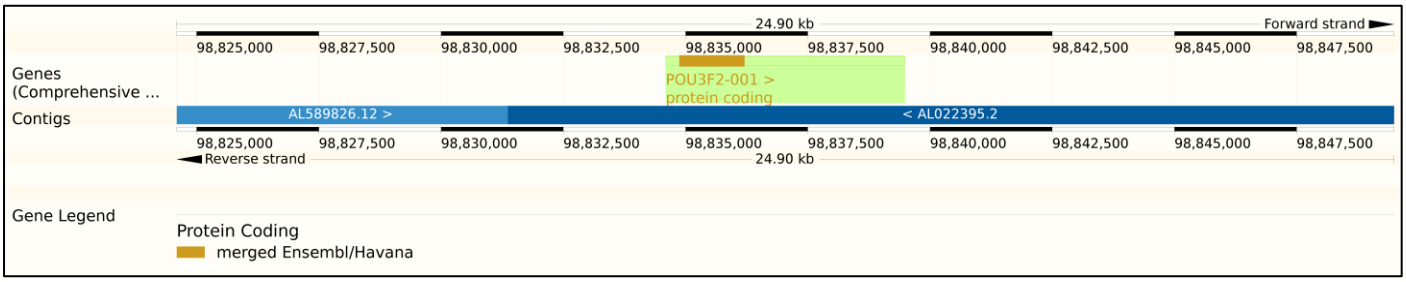
Figure S9: The evolutionary syntenic architecture of the 6q16.2 region shows that POU3F2 and SIM1 are not located on the same chromosome in a number of mammals.



Snapshot from Genomicus genome browser. The figure is centred on POU3F2 gene which is indicated with a red box and arrow. SIM1 location and orientation is indicated by blue box and arrow. The right column gives the name of the species. Faded boxes indicate that the genes are not syntenically linked in the respective specie. Mouse Pou3f2 and Sim1 are highlighted demonstrating that unlike in humans, these two genes are on different chromosomes.

Figure S10: ***POU3F2 is a single exon gene with a number of repeats and thus challenging to sequence***

A



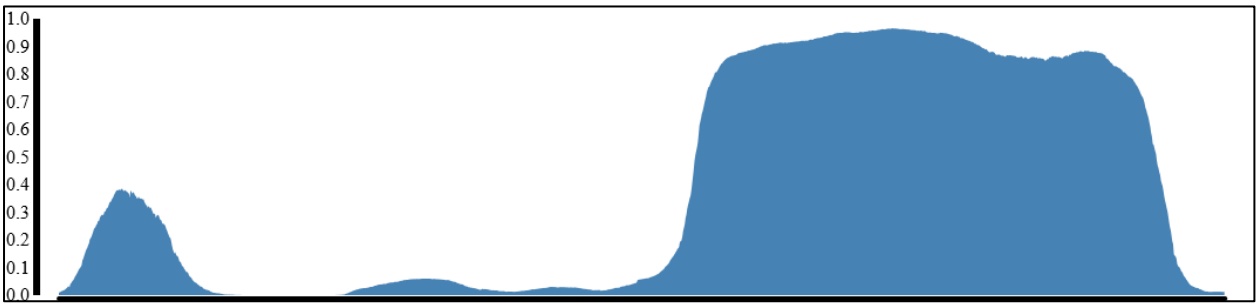
B

```
 241  GCGGCTCCTTTAACCGGAGCGCTCAGTCCGGCTCCGAGAGTCATGGCGACCGCAGCGTCT
      .........................................ATGGCGACCGCAGCGTCT
      .........................................-M--A--T--A--A--S-

 301  AACCACTACAGCCTGCTCACCTCCAGCGCCTCCATCGTGCACGCCGAGCCGCCCGGCGGC
  19  AACCACTACAGCCTGCTCACCTCCAGCGCCTCCATCGTGCACGCCGAGCCGCCCGGCGGC
   7  -N--H--Y--S--L--L--T--S--S--A--S--I--V--H--A--E--P--P--G--G-

 361  ATGCAGCAGGGCGCGGGGGGCTACCGCGAAGCGCAGAGCCTGGTGCAGGGCGACTACGGC
  79  ATGCAGCAGGGCGCGGGGGGCTACCGCGAAGCGCAGAGCCTGGTGCAGGGCGACTACGGC
  27  -M--Q--Q--G--A--G--G--Y--R--E--A--Q--S--L--V--Q--G--D--Y--G-

 421  GCTCTGCAGAGCAACGGACACCCGCTCAGCCACGCTCACCAGTGGATCACCGCGCTGTCC
 139  GCTCTGCAGAGCAACGGACACCCGCTCAGCCACGCTCACCAGTGGATCACCGCGCTGTCC
  47  -A--L--Q--S--N--G--H--P--L--S--H--A--H--Q--W--I--T--A--L--S-

 481  CACGGCGGCGGCGGCGGGGGCGGTGGCGGCGGCGGGGGGGGCGGGGGCGGCGGCGGGGGC
 199  CACGGCGGCGGCGGCGGGGGCGGTGGCGGCGGCGGGGGGGGCGGGGGCGGCGGCGGGGGC
  67  -H--G--G--G--G--G--G--G--G--G--G--G--G--G--G--G--G--G--G--G-

 541  GGCGGCGACGGCTCCCCGTGGTCCACCAGCCCCCTGGGCCAGCCGGACATCAAGCCCTCG
 259  GGCGGCGACGGCTCCCCGTGGTCCACCAGCCCCCTGGGCCAGCCGGACATCAAGCCCTCG
  87  -G--G--D--G--S--P--W--S--T--S--P--L--G--Q--P--D--I--K--P--S-

 601  GTGGTGGTGCAGCAGGGCGGCCGCGGAGACGAGCTGCACGGGCCAGGCGCCCTGCAGCAG
 319  GTGGTGGTGCAGCAGGGCGGCCGCGGAGACGAGCTGCACGGGCCAGGCGCCCTGCAGCAG
 107  -V--V--V--Q--Q--G--G--R--G--D--E--L--H--G--P--G--A--L--Q--Q-

 661  CAGCATCAGCAGCAGCAACAGCAACAGCAGCAGCAACAGCAGCAACAGCAGCAGCAGCAG
 379  CAGCATCAGCAGCAGCAACAGCAACAGCAGCAGCAACAGCAGCAACAGCAGCAGCAGCAG
 127  -Q--H--Q--Q--Q--Q--Q--Q--Q--Q--Q--Q--Q--Q--Q--Q--Q--Q--Q--Q-

 721  CAGCAACAGCGGCCGCCGCATCTGGTGCACCACGCCGCTAACCACCACCCGGGACCCGGG
 439  CAGCAACAGCGGCCGCCGCATCTGGTGCACCACGCCGCTAACCACCACCCGGGACCCGGG
 147  -Q--Q--Q--R--P--P--H--L--V--H--H--A--A--N--H--H--P--G--P--G-

 781  GCATGGCGGAGCGCGGCGGCTGCAGCGCACCTCCCACCCTCCATGGGAGCGTCCAACGGC
 499  GCATGGCGGAGCGCGGCGGCTGCAGCGCACCTCCCACCCTCCATGGGAGCGTCCAACGGC
 167  -A--W--R--S--A--A--A--A--A--H--L--P--P--S--M--G--A--S--N--G-

 841  GGCTTGCTCTACTCGCAGCCCAGCTTCACGGTGAACGGCATGCTGGGCGCCGGCGGGCAG
 559  GGCTTGCTCTACTCGCAGCCCAGCTTCACGGTGAACGGCATGCTGGGCGCCGGCGGGCAG
 187  -G--L--L--Y--S--Q--P--S--F--T--V--N--G--M--L--G--A--G--G--Q-

 901  CCGGCCGGTCTGCACCACCACGGCCTGCGGGACGCGCACGACGAGCCACACCATGCCGAC
 619  CCGGCCGGTCTGCACCACCACGGCCTGCGGGACGCGCACGACGAGCCACACCATGCCGAC
 207  -P--A--G--L--H--H--H--G--L--R--D--A--H--D--E--P--H--H--A--D-

 961  CACCACCCGCACCCGCACTCGCACCCACACCAGCAGCCGCCGCCCCCGCCGCCCCCGCAG
 679  CACCACCCGCACCCGCACTCGCACCCACACCAGCAGCCGCCGCCCCCGCCGCCCCCGCAG
 227  -H--H--P--H--P--H--S--H--P--H--Q--Q--P--P--P--P--P--P--P--Q-

1021  GGTCCGCCTGGCCACCCAGGCGCGCACCACGACCCGCACTCGGACGAGGACACGCCGACC
 739  GGTCCGCCTGGCCACCCAGGCGCGCACCACGACCCGCACTCGGACGAGGACACGCCGACC
 247  -G--P--P--G--H--P--G--A--H--H--D--P--H--S--D--E--D--T--P--T-
```

```
1081 TCGGACGACCTGGAGCAGTTCGCCAAGCAGTTCAAGCAGCGGCGGATCAAACTGGGATTT
 799 TCGGACGACCTGGAGCAGTTCGCCAAGCAGTTCAAGCAGCGGCGGATCAAACTGGGATTT
 267 -S--D--D--L--E--Q--F--A--K--Q--F--K--Q--R--R--I--K--L--G--F-

1141 ACCCAAGCGGACGTGGGGCTGGCTCTGGGCACCCTGTATGGCAACGTGTTCTCGCAGACC
 859 ACCCAAGCGGACGTGGGGCTGGCTCTGGGCACCCTGTATGGCAACGTGTTCTCGCAGACC
 287 -T--Q--A--D--V--G--L--A--L--G--T--L--Y--G--N--V--F--S--Q--T-

1201 ACCATCTGCAGGTTTGAGGCCCTGCAGCTGAGCTTCAAGAACATGTGCAAGCTGAAGCCT
 919 ACCATCTGCAGGTTTGAGGCCCTGCAGCTGAGCTTCAAGAACATGTGCAAGCTGAAGCCT
 307 -T--I--C--R--F--E--A--L--Q--L--S--F--K--N--M--C--K--L--K--P-

1261 TTGTTGAACAAGTGGTTGGAGGAGGCGGACTCGTCCTCGGGCAGCCCCACGAGCATAGAC
 979 TTGTTGAACAAGTGGTTGGAGGAGGCGGACTCGTCCTCGGGCAGCCCCACGAGCATAGAC
 327 -L--L--N--K--W--L--E--E--A--D--S--S--S--G--S--P--T--S--I--D-

1321 AAGATCGCAGCGCAAGGGCGCAAGCGGAAAAAGCGGACCTCCATCGAGGTGAGCGTCAAG
1039 AAGATCGCAGCGCAAGGGCGCAAGCGGAAAAAGCGGACCTCCATCGAGGTGAGCGTCAAG
 347 -K--I--A--A--Q--G--R--K--R--K--K--R--T--S--I--E--V--S--V--K-

1381 GGGGCTCTGGAGAGCCATTTCCTCAAATGCCCCAAGCCCTCGGCCCAGGAGATCACCTCC
1099 GGGGCTCTGGAGAGCCATTTCCTCAAATGCCCCAAGCCCTCGGCCCAGGAGATCACCTCC
 367 -G--A--L--E--S--H--F--L--K--C--P--K--P--S--A--Q--E--I--T--S-

1441 CTCGCGGACAGCTTACAGCTGGAGAAGGAGGTGGTGAGAGTTTGGTTTTGTAACAGGAGA
1159 CTCGCGGACAGCTTACAGCTGGAGAAGGAGGTGGTGAGAGTTTGGTTTTGTAACAGGAGA
 387 -L--A--D--S--L--Q--L--E--K--E--V--V--R--V--W--F--C--N--R--R-

1501 CAGAAAGAGAAAAGGATGACCCCTCCCGGAGGGACTCTGCCGGGCGCCGAGGATGTGTAC
1219 CAGAAAGAGAAAAGGATGACCCCTCCCGGAGGGACTCTGCCGGGCGCCGAGGATGTGTAC
 407 -Q--K--E--K--R--M--T--P--P--G--G--T--L--P--G--A--E--D--V--Y-

1561 GGGGGGAGTAGGGACACTCCACCACACCACGGGGTGCAGACGCCCGTCCAGTGAACTCGA
1279 GGGGGGAGTAGGGACACTCCACCACACCACGGGGTGCAGACGCCCGTCCAGTGA......
 427 -G--G--S--R--D--T--P--P--H--H--G--V--Q--T--P--V--Q--*-......
```

C

D



A. *Snapshot from Ensembl genome browser from Gene Summary page for POU3F2 (ENSG00000184486) demonstrating that this gene has only one exon.*
B. *POU3F2 cDNA sequence. The long repeats in the DNA sequence are highlighted.*
C. *Domains of POU3F2 protein*
D. *ExAC coverage plot for POU3F2. X-axis denotes nucleotide positions (unnumbered) in POU3F2 exon and y-axis number of individuals with more than 30X coverage for the corresponding nucleotide position on the X-axis.*

Table S1: Monogenic obesity syndromes in the Leptin>melanocortin>SIM1 neuro-endocrine pathway.

Gene	Phenotype	OMIM	Reference
LEP	Morbid obesity due to leptin deficiency	614962	[5]
LEPR	Morbid obesity due to leptin receptor deficiency	614963	[6,7]
POMC	Obesity, adrenal insufficiency and red hair syndrome	609734	[8]
PCSK1	Obesity with impaired prohormone processing	600955	[9]
MC4R	Autosomal dominant obesity	601665	[10,11]
MRAP2	Susceptibility to obesity BMIQ18	615457	[12]
SIM1	Severe obesity	601665	[13–15]

Table S2: List of genes within the deletion in Family 1 and properties of their variation.

Table based on data accessed from http://exac.broadinstitute.org/ on 15th September 2015.

Gene	Number of potentially truncating variants	Combined allele frequency of truncating variants	HI index	Associated disease (OMIM)	Inheritance pattern of the disease
POU3F2	0	0	15.18	None	Not applicable
FBXL4	11	$3.1x10^{-4}$	10.28	Encephalomyopathic type mitochondrial DNA depletion syndrome (605654)	Autosomal recessive
FAXC	2	$1.7x10^{-5}$	21.43	None	Not applicable
COQ3	19	$8.2x10^{-4}$	40.90	None	Not applicable
PNISR	11	$2.2\ x10^{-4}$	9.31	None	Not applicable
USP45	57	$1.9X10^{-2}$	45.26	None	Not applicable
TSTD3	No data	No data	No data	None	Not applicable
CCNC	10	$3.8x10^{-1}$	2.44	None	Not applicable
PRDM13	8	$8.1x10^{5}$	56.20	None	Not applicable

Table S3: Extent of 6q16 deletion in all the families studied in this project.

Family #	Start co-ordinate (Hg19)	End co-ordinate (Hg19)	Size (Mb)
1	chr6:99218535	chr6:100260996	1.04
2	chr6:98583798	chr6:99582590	0.99
3	chr6:98981807	chr6:100377741	1.39
4	chr6:98981812	chr6:100261128	1.28
5	chr6:99179231	chr6:100641841	1.46
6	chr6:99218553	chr6:100319568	1.10

Table S4: Amino acid identities in the amino terminal region of POU3 family proteins.

Outside of the homeodomain there is 59% identity between Pou3f2a and Pou3f2b, and 50% identity between POU3F2 and Pou3f2a or Pou3f2b. The amino acid identity among Pou3f2a/b and other members of the POU family is reduced in comparison.

POU family protein	Pou3f2a	Pou3f2b
zf_Pou3f2b	59%	-
Hu_POU3F1	35%	30%
Hu_POU3F2	50%	50%
Hu_POU3F3	42%	28%
Hu_POU3F4	47%	45%

Table S5: pou3f2a/b MO sequences.

pou3f2a 5’	AGTGGTTGGACGCCGCGGTCGCCAT
control-pou3f2a 5’	AGTcGTTGcACcCCGCcGTCcCCAT
pou3f2b 5'	GATTGGATGCTGTAGTCGCCATGAC
control-pou3f2b 5’	GAaTGcATGCTGTAGTCGCgATcAC

The underlined nucleotides mark the AUG translation initiation site, while the lower case letters in the control MO indicates the bases that are mis-matched.

Table S6: Primers and reaction conditions for arnt2^{hi2639cTg} genotyping

Hi2639_3E01 (F1):	5’ ATA CTG AGG GTG AAC GCA GAC G 3’
Hi2639_3E02 (R1):	5’ TCG CTT CTC GCT TCT GTT CG 3’
arnt2 Endogenous:	5’ AAA TCG CAT CCA AGC ATC GC 3’

Embryos were genotyped for the arnt2^{hi2639cTg} allele by tail clip. DNA was prepared by alkaline lysis (25 mM NaOH, 0.2 mM disodium EDTA) for 1 hour at 95° C, followed by neutralization with an equal amount of 40 mM Tris-HCl; pH5. Amplification conditions were as per ZIRC PCR protocol (http://zebrafish.org/zirc/fish/pdf/pcr/hi2639cTg.pdf), with the addition of a third primer to identify the endogenous allele. The PCR reaction was performed using 1X PCR Master Mix (Thermo Scientific), 0.4 uM F1 primer, 0.2 uM R1 primer, 0.2 uM Endogenous primer, and 4 µl of tail clip DNA, in a 25 µl reaction solution. The *arnt2* PCR program was run according to the following protocol: 94°C, 3 minutes; 94°C, 30 seconds; 62°C, 40 seconds; 72°C, 30 seconds; repeat steps 2-4, 34 cycles, 72°C, 5 minute extention. The entire reaction was run on a 2% agarose gel, producing a mutant band (arnt2^{hi2639cTg} = 342 bp), a wild-type band (wild-type = 237 bp) or both for the heterozygous genotype.

Supplementary References

1. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478,** 483–489 (2011).
2. Hawrylycz, M. *et al.* in *Springer Handbook of Bio-/Neuroinformatics* (ed. Kasabov, N.) 1111–1126 (Springer Berlin Heidelberg, 2014).
3. Sugitani, Y. *et al.* Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons. *Genes Dev.* **16,** 1760–1765 (2002).
4. Belot, A. *et al.* Protein Kinase Cδ Deficiency Causes Mendelian Systemic Lupus Erythematosus With B Cell-Defective Apoptosis and Hyperproliferation. *Arthritis Rheum.* **65,** 2161–2171 (2013).
5. Farooqi, I. S. *et al.* Partial leptin deficiency and human adiposity. *Nature* **414,** 34–35 (2001).
6. Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392,** 398–401 (1998).
7. Farooqi, I. S. *et al.* Clinical and Molecular Genetic Spectrum of Congenital Deficiency of the Leptin Receptor. *N. Engl. J. Med.* **356,** 237–247 (2007).
8. Krude, H. *et al.* Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat. Genet.* **19,** 155–157 (1998).
9. Jackson, R. S. *et al.* Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat. Genet.* **16,** 303–306 (1997).
10. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat. Genet.* **20,** 113–114 (1998).
11. Farooqi, I. S. *et al.* Clinical Spectrum of Obesity and Mutations in the Melanocortin 4 Receptor Gene. *N. Engl. J. Med.* **348,** 1085–1095 (2003).
12. Asai, M. *et al.* Loss of Function of the Melanocortin 2 Receptor Accessory Protein 2 Is Associated with Mammalian Obesity. *Science* **341,** 275–278 (2013).

13. Holder, J. L., Butte, N. F. & Zinn, A. R. Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. *Hum. Mol. Genet.* **9,** 101–108 (2000).
14. Bonnefond, A. *et al.* Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi–like features. *J. Clin. Invest.* **123,** 3037–3041 (2013).
15. Ramachandrappa, S. *et al.* Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *J. Clin. Invest.* **123,** 3042–3050 (2013).