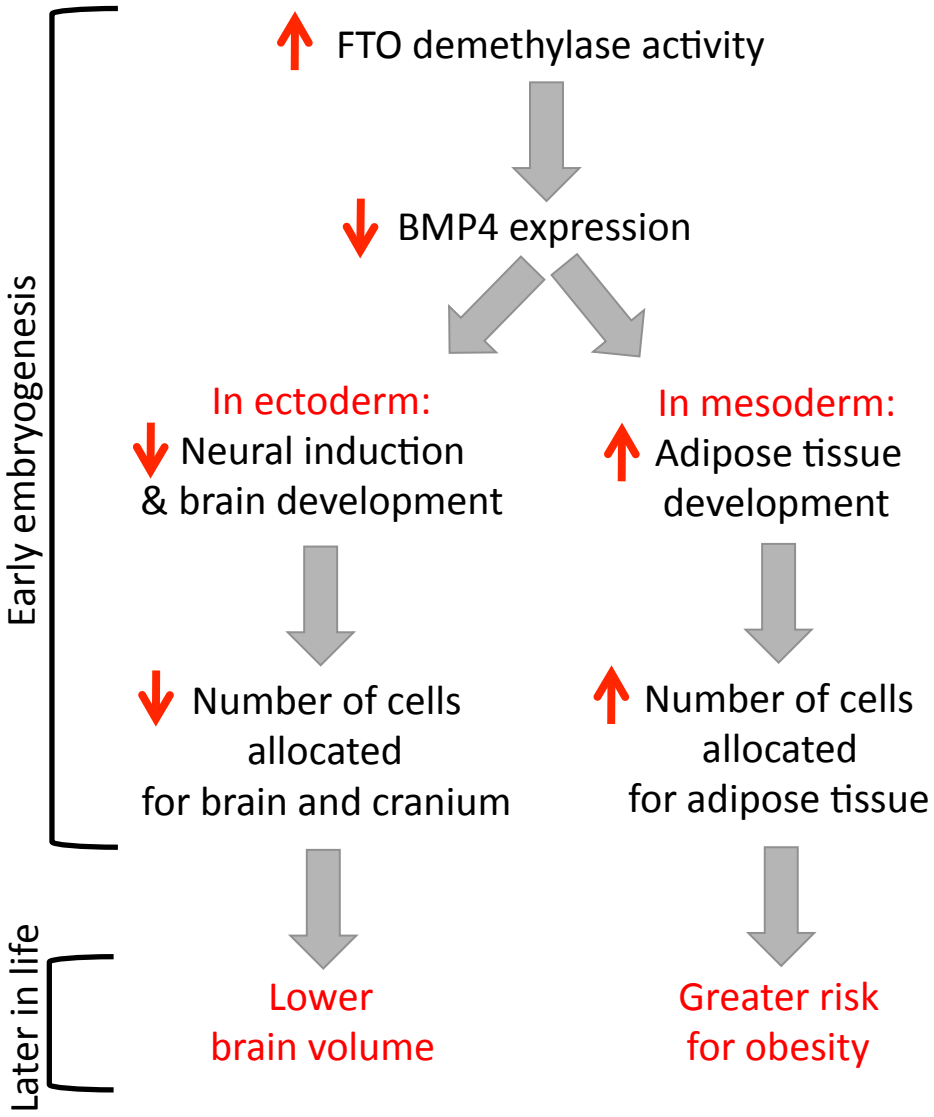
SUPPLEMENTARY INFORMATION

Supplementary Figure 1: A scenario of *FTO* inverse effects on adipose and brain tissues. BMP4 (bone morphogenetic protein 4) is a protein involved in the regulation of cell proliferation and differentiation during embryonic development; its inhibition in ectoderm is critical for neural induction and the development of the brain, whereas its inhibition in mesoderm may compromise the development of adipose tissue. We speculate that a moderate increase in *BMP4* expression, due to greater demethylating activity of FTO, would lead to a decrease in allocating stem cells towards brain-cell lineages and an increase in allocating stem cells towards adipose-tissue lineages. Such relative differences in stem-cell allocation, and thus life-long potential for growth of the respective tissues, would then result in a lower brain volume and a higher body-fat mass postnatally. The former would be realized during major periods of brain development (prenatal and early postnatal period), whereas the latter would become apparent gradually during postnatal life as a result of a chronic positive energy balance (life-long risk for obesity).

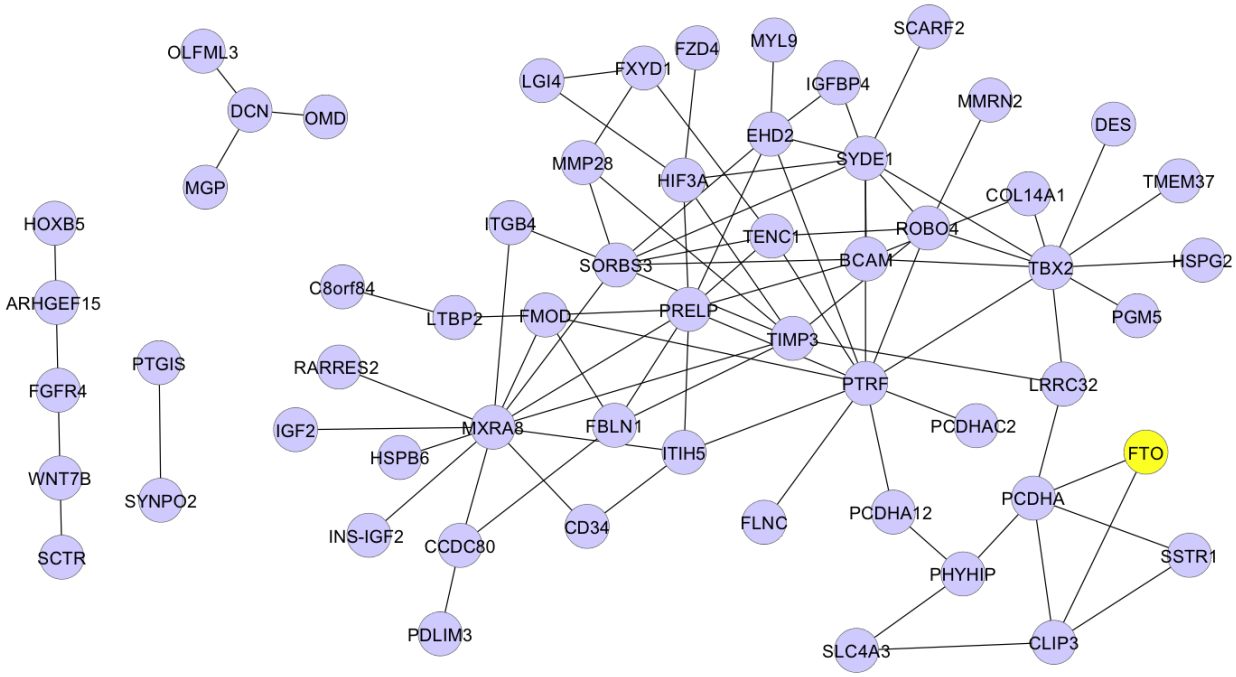
Supplementary Figure 2: A subnetwork of genes prioritized for embryonic involvement with *FTO*. *FTO* and *BMP4* exhibited a significantly high co-expression in embryonic co-expression datasets. We identified 165 genes showing a similar correlation profile with *FTO* across datasets as *FTO* did with *BMP4* (significantly correlated; $r>0.25$) and examined the role of these genes within a co-expression network constructed from brain-derived expression data. Links indicate co-expression in the top 0.5% of values among all gene-gene pairs. These genes show significantly low external connectivity and significantly high internal connectivity. We note that *FTO* is particularly strongly co-expressed with the proto-cadherin alpha gene cluster (PCDHA). Even though these 165 genes were identified on the basis of showing a similar correlation profile with *FTO* across datasets as *FTO* did with *BMP4*, *FTO* and *BMP4* did not themselves exhibit significant co-expression within this sub-network and *BMP4* was not co-expressed strongly within this sub-network.

Supplementary Figure 3: Brain volume, intra-cranial volume and height: pair-wise comparisons. Data from the SYS-Discovery sample were adjusted for potentially confounding effects of age and sex.

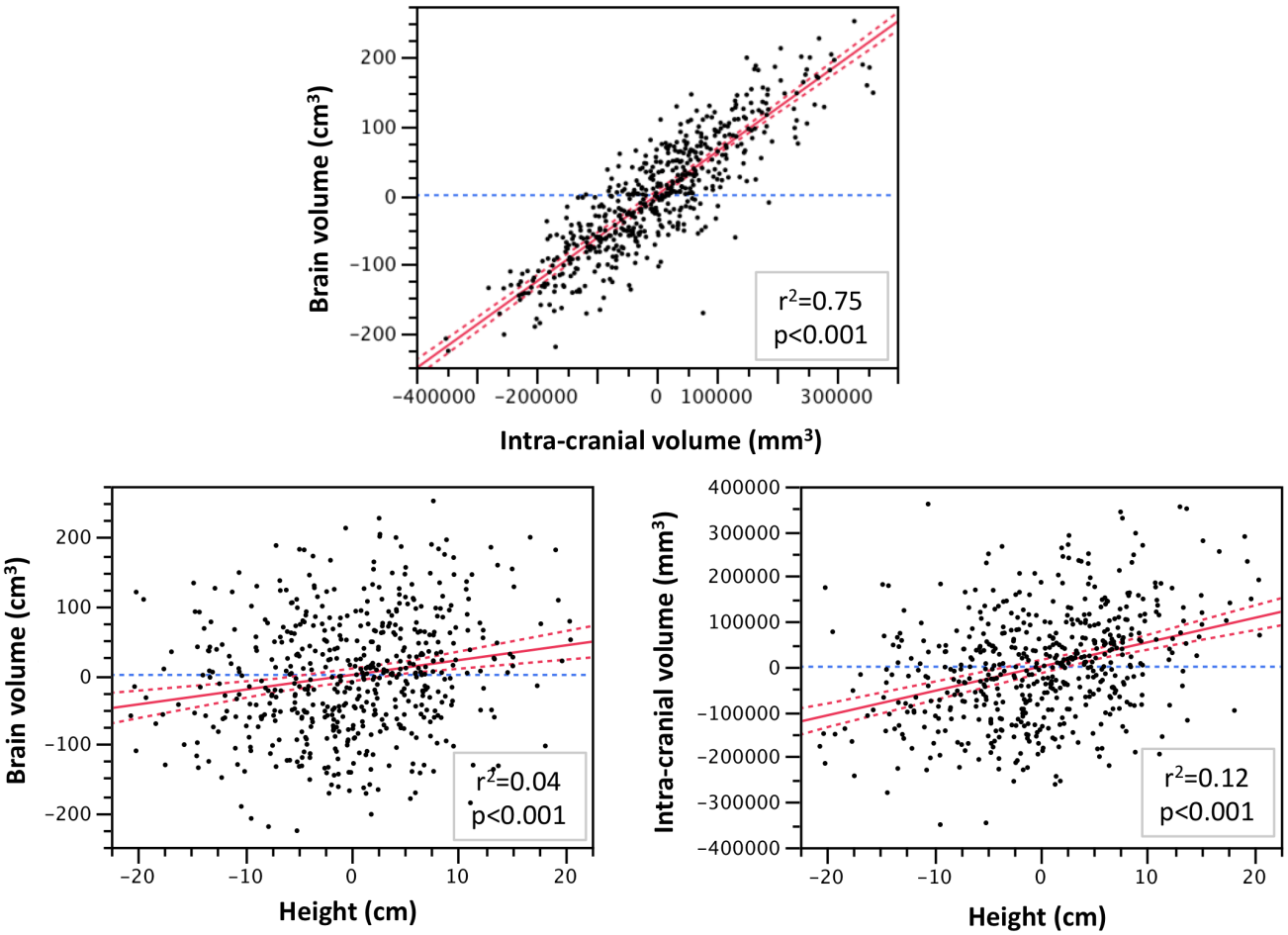
Supplementary Figure 4: Intra-cranial mask. From left to right: Sagittal, coronal and axial views of a population average derived from the SYS images and the manually defined mask used to estimate intra-cranial volume.



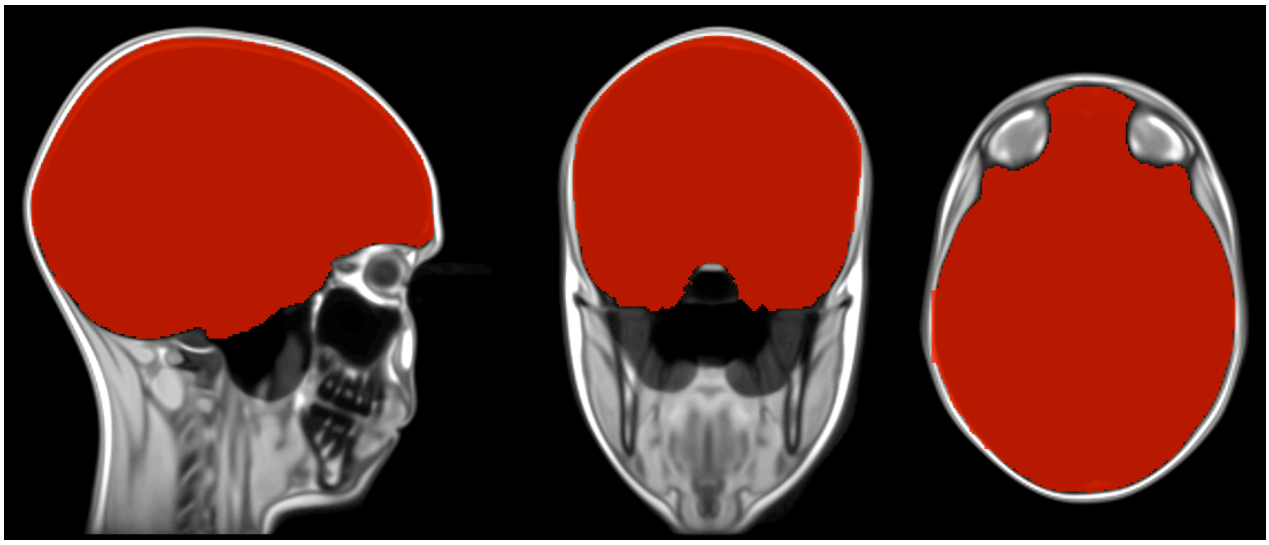
Supplementary Figure 1.



Supplementary Figure 2.



Supplementary Figure 3.



Supplementary Figure 4.

Supplementary Table 1A: Correlation matrix of PCA variables in the SYS-Discovery sample

	Brain volume	TBF	LBM	Height
Brain volume	1	0.06	0.15	0.21
TBF	0.06	1	0.47	0.22
LBM	0.15	0.47	1	0.71
Height	0.21	0.22	0.71	1

Supplementary Table 1B: Correlation matrix of PCA variables in the SYS-Replication sample

	Brain volume	TBF	LBM	Height
Brain volume	1	0.15	0.26	0.28
TBF	0.15	1	0.60	0.26
LBM	0.26	0.60	1	0.70
Height	0.28	0.26	0.70	1

Supplementary Table 1C: Correlation matrix of PCA variables in the IMAGEN sample

	Brain volume	Body weight	Height
Brain volume	1	0.11	0.22
Body weight	0.11	1	0.55
Height	0.22	0.55	1

Pair-wise correlation coefficients are shown; all variables were adjusted for age and sex.
PCA: principal component analysis, TBF: total body fat, LBM: lean body mass.

Supplementary Table 2: Associations between *FTO* (rs9930333) and body fat, brain volume and shared inverse variance between body fat and brain volume in the SYS-Discovery – additional confounders

Outcome	Effect size±SE (G-allele)		P-value	
	Model 1	Model 2	Model 1	Model 2
TBF (log kg)	0.052±0.016	0.04±0.02	0.002	0.01
Brain volume (cm^3)	-16.8±6.0	-21.4±6.4	0.006	0.001
Component 2	-0.29±0.06	-0.29±0.08	$9.0x10^{-6}$	$1.2x10^{-4}$

Associations were tested with Merlin-1.1.2, while adjusting for:
SE: standard error
Model 1: age and sex
Model 2: age, sex, annual family income, maternal education, exercise (number of at least 20-min sessions of strenuous exercise per week), and energy intake (24-hour food recall). Details of these additional assessments are described in (30).
TBF: total body fat; Component 2: principal component 1 from principal component analysis (PCA) of brain volume and BMI determinants (i.e. total body fat, lean body mass and height)

Meta-analysis of *FTO* expression profile

Rationale

Co-expression is a noisy signature for co-functionality and may be misleading if the shared functionality of two genes is very specific and not co-expressed in the majority of tissues or under the majority of conditions. On the other hand, the presence of this specificity in co-expression data may be advantageous if we wish to probe whether two genes (or their products) interact only under a subset of conditions. To determine whether there was any particular functional profile to *FTO*'s possible interaction with *BMP4* we conducted a two-stage analysis.

In the first stage, a large set of annotated experiments was assembled and, in each one, the co-expression of *FTO* with *BMP4* was assessed relative to other genes. If, for example, *FTO* and *BMP4* were co-expressed preferentially (and significantly) in experiments involving brain tissue, we would infer a joint role that is somewhat specific to the brain. This relies on variability in the co-expression between *FTO* and *BMP4*; if they were perfectly co-expressed (e.g., a complex-like interaction) across all experiments, then there would be no functional preference. If the interaction were even somewhat specific, we would not even expect *FTO* and *BMP4* to be co-expressed in the aggregate of all experiments. Methodologically, this almost exactly resembles enrichment analysis, except instead of a ranked list of genes (with corresponding functional annotations), we had a ranked list of experiments (with corresponding functional annotations). We then identified a list of 165 genes that had the same profile across experiments as *BMP4* had with *FTO*. Note that this does not mean these genes will tend to be co-expressed on average.

In our second stage of analysis, our intent was to determine if this list of genes could be seen to have a specific co-expression role. In particular, we investigated whether this set of genes was co-expressed (or exhibited significant modularity) in brain co-expression data vs. non-brain co-expression data in previously constructed networks.

Methods

For analysis of *FTO*'s expression profile, we assembled 721 publicly available expression experiments constituting 34,019 individual microarrays from the Gemma system[1] and conducted co-expression analysis as described in [2]. Briefly, correlations between *FTO*'s expression profile and profiles of each of 14,184 genes that were present in at least 500 expression data sets (each using many individual microarrays) were analyzed for each data set. These correlations were then replaced with ranks, to give an expression profile similarity score between *FTO* and each other gene in each data set. Next, each data set was annotated by 1 to 128 terms (such as tissue type; only terms present in at least 5 [up to 25] experiments were considered). This allowed us to examine if *FTO* exhibited changes in the similarity of its expression profile with other genes depending on experimental annotation. This closely resembles gene-function enrichment analysis (e.g. [3]) and we applied the same methodology to test for enrichment across data set annotations.

Some further details are available at http://www.chibi.ubc.ca/FTO.

Results and Discussion

Using our experimental annotations and the *FTO*-*BMP4*'s co-expression score as a ranking tool, we observed that data sets with the "*Adipose Tissue*" annotation were the most significantly enriched. ($p<0.01$, see supplemental website for experiment lists). The other comparably enriched annotation ($p<0.01$) was "*Behavioral Activity*", indicating that the interaction between *FTO* and *BMP4* may vary between adipose tissue and behavioral activity (including neurological function).

In order to capture cases where *FTO* and *BMP4* may be involved with strong but divergent effects, we tested for enrichment of the ranks of the absolute co-expression values. In this case "*Embryonic*" data sets exhibited the most significant change in co-expression similarity ($p<0.01$).

We next examined the set of genes whose interaction strength with *FTO* was significantly

correlated with *BMP4*'s correlation strength with *FTO* ($r>0.25$, $p<0.01$). This identified a set of 165 genes that was significantly enriched for "*Nervous System Development*" within its top 10 Gene Ontology terms (corrected $p<0.01$). This subnetwork of genes exhibited significant modularity ($p<0.01$, permutation test) in a network constructed from the brain-derived expression data but not from non-brain-derived expression data (size and platform matched). *FTO* exhibited particularly strong interaction with the proto-cadherin alpha gene cluster (see Supplementary Figure 1). Even though this set of 165 genes was identified on the basis of similarity with *BMP4-FTO*'s correlation strength, *FTO* and *BMP4* did not themselves exhibit significant co-expression within this sub-network and *BMP4* was not co-expressed strongly within this data.

References

1 Zoubarev, A. *et al.* Gemma: A resource for the re-use, sharing and meta-analysis of expression profiling data. *Bioinformatics*, doi:10.1093/bioinformatics/bts430 (2012).

2 Gillis, J. & Pavlidis, P. The Role of Indirect Connections in Gene Networks in Predicting Function. *Bioinformatics* (2011).

3 Gillis, J., Mistry, M. & Pavlidis, P. Gene function analysis in complex data sets using ErmineJ. *Nature Protocols* **5**, 1148-1159 (2010).