Genomics Data 12 (2017) 24–27

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Hypothalamic transcriptome analysis of congenic-derived F_2 mice (chromosome 17:3–45 Mb) exhibiting preferential carbohydrate (versus fat) intake

Brenda K. Richards *, Claudia Kappen, Susan Newman

Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, LA 70808-4124, USA

ARTICLE INFO

Article history:
Received 31 January 2017
Accepted 15 February 2017
Available online 24 February 2017

Specifications	
Organism/cell line/tissue	*Mus musculus*, Whole hypothalamus from B6·CAST-17.1 congenic-derived F2 mice, possessing a CAST donor segment (Chr 17:3.19–45.73 Mb) genotype of *cast*/*cast* (CAST/EiJ) or *b6*/*b6* (C57BL/6J) on an otherwise B6 genome.
Sex	Male
Sequencer or array type	5500XL SOLiD next generation sequencer (Life Technologies)
Data format	Raw (.csfasta) and processed (.txt)
Experimental factors	2 genotypes and food intake in kilocalorie (kcal) from the carbohydrate-rich diet in a diet choice paradigm (see below).
Experimental features	For 2 d, mice were provided a choice between two diet mixtures containing either 78% fat/22% protein or 78% carbohydrate/22% protein (per cent by energy) and then were euthanized for collection of whole hypothalamus. Before the 2 d test, animals completed a 10 d macronutrient diet selection test, followed by a two week wash-out on standard chow. Experimental animals used for tissue collection were chosen from a larger group based on self-selected intake of carbohydrate + protein kcal in the 10 d diet choice test: congenic *cast*/*cast* F2s from the upper quartile ($n = 12$) and congenic *b6*/*b6* F2 s from the lowest quartile ($n = 12$).
Consent	N/A
Sample source location	Baton Rouge, LA USA

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60756.

2. Experimental design, materials and methods

2.1. Experimental design

Along with the general drive to eat, macronutrient-specific appetites are likely encoded by unique molecular changes in the hypothalamus, which occur largely as the result of unknown interoceptive signaling [1]. We conducted a transcriptome analysis of total RNA extracted from the whole hypothalamus of B6·CAST-17.1 congenic F2 mice that were either homozygous *cast*/*cast* or *b6*/*b6* in the chromosome 17 congenic interval (Chr 17:3.19–45.73 Mb), on an otherwise B6 genome (see strain description below). A complete description of samples used and files that are available on GEO is found in Table 1.

2.2. Mice

Our development of the B6·CAST-17.1 congenic strain has been described previously [2–4]. Here, the 42.5 Mb *cast*/*cast* congenic donor segment was delimited by SNP markers *rs49640908* (proximal) and *rs48762654* (distal). To fine-map the *Mnic1* QTL using genetic recombination, a large congenic-by-recipient F_2 population was generated [5]. For transcriptome analysis, only non-recombinant congenic F_2 littermates were selected, i.e., those possessing a

* Corresponding author.
E-mail address: richarbk@pbrc.edu (B.K. Richards).

http://dx.doi.org/10.1016/j.gdata.2017.02.009

Table 1
Sample information.

Sample name	Raw/qual	Processed	Total reads	Total mapped reads	%Mapped	Unique mapped	%Unique mapped	Multi-position mapped	%Multi-position mapped
B6_24278_F3	.csfasta	.txt	23,361,768	10,178,992	44%	8,885,889	38%	1,293,103	6%
B6_24541_F3	.csfasta	.txt	19,694,925	8,759,552	44%	7,769,217	39%	990,335	5%
B6_24550_F3	.csfasta	.txt	27,226,120	11,100,128	41%	9,562,805	35%	1,537,323	6%
B6_24574_F3	.csfasta	.txt	25,833,763	11,486,371	44%	9,990,958	39%	1,495,413	6%
B6_25014_F3	.csfasta	.txt	27,188,757	11,700,353	43%	10,091,604	37%	1,608,749	6%
B6_25874_F3	.csfasta	.txt	26,442,221	11,800,576	45%	10,278,749	39%	1,521,827	6%
B6_25907_F3	.csfasta	.txt	30,273,294	13,644,050	45%	12,182,361	40%	1,461,689	5%
B6_26480_F3	.csfasta	.txt	26,768,937	11,065,979	41%	9,720,578	36%	1,345,401	5%
B6_26510_F3	.csfasta	.txt	21,626,926	9,592,296	44%	8,548,103	40%	1,044,193	5%
B6_26545_F3	.csfasta	.txt	27,962,570	12,754,249	46%	11,349,041	41%	1,405,208	5%
B6_26546_F3	.csfasta	.txt	26,349,724	11,470,527	44%	10,155,815	39%	1,314,712	5%
B6_26641_F3	.csfasta	.txt	31,913,005	14,502,508	45%	13,092,465	41%	1,410,043	4%
CAST_24297_F3	.csfasta	.txt	24,744,288	10,414,231	42%	9,242,888	37%	1,171,343	5%
CAST_24596_F3	.csfasta	.txt	27,570,206	11,154,300	40%	9,886,768	36%	1,267,532	5%
CAST_25089_F3	.csfasta	.txt	29,151,097	11,939,297	41%	9,886,749	34%	2,052,548	7%
CAST_25105_F3	.csfasta	.txt	27,766,526	11,562,831	42%	10,019,642	36%	1,543,189	6%
CAST_25127_F3	.csfasta	.txt	27,268,698	12,524,167	46%	10,794,182	40%	1,729,985	6%
CAST_25157_F3	.csfasta	.txt	24,714,864	11,310,343	46%	9,729,647	39%	1,580,696	6%
CAST_25159_F3	.csfasta	.txt	29,730,733	12,776,479	43%	11,001,485	37%	1,774,994	6%
CAST_25489_F3	.csfasta	.txt	30,248,262	12,565,504	42%	10,767,861	36%	1,797,643	6%
CAST_25932_F3	.csfasta	.txt	28,354,944	13,507,960	48%	11,912,291	42%	1,595,669	6%
CAST_25992_F3	.csfasta	.txt	28,947,713	13,353,512	46%	11,221,898	39%	2,131,614	7%
CAST_25994_F3	.csfasta	.txt	34,699,812	13,657,150	39%	12,085,839	35%	1,571,311	5%
CAST_26001_F3	.csfasta	.txt	25,010,571	10,742,796	43%	8,885,889	37%	1,426,425	6%

The table lists the file types available in GEO (.csfasta and .txt) as well as the total number of reads for each sample generated using the 5500XL SOLiD system (Life Technologies). Only perfectly matched sequences were counted.

homozygous genotype of either *cast*/*cast* or *b6*/*b6* throughout the 42.5 Mb chromosome 17 congenic region, and a *b6*/*b6* genotype across the rest of the genome. All animal procedures were approved by the Pennington Biomedical Research Center Institutional Animal Care and Use Committee.

2.3. Animal genotyping

Genomic DNA was isolated from spleen and a custom SNP panel (Illumina)was designed to genotype the animals; for details, see [5]. Genotypes were coded for the chromosome 17 congenic segment based on 301 SNP markers.

2.4. Phenotyping

Food intake was determined in singly-housed adult male mice using a two-choice macronutrient diet protocol. For 10 days mice were given a choice between two diets: a fat + protein versus a carbohydrate + protein mixture, each containing protein (casein) (22% of energy) with the balance of calories contributed by either fat (vegetable shortening) or carbohydrate (corn starch-sucrose) (78% of energy). Both diets contained minerals, vitamins and cellulose. Diet intake including all spillage were measured daily to 0.1 g. The experimental diet composition and complete details of the phenotyping procedures have been previously described [5,6].

2.5. Sample selection for gene expression analyses

This experiment was carried out using B6·CAST-17.1 congenic-derived F2 mice that possessed a non-recombined, chromosome 17 donor interval (3.19–45.73 Mb). Based on high-density SNP marker analysis, we determined the absence of recombination in this congenic interval, resulting in genotypes of either *cast*/*cast* ($n = 44$) or *b6*/*b6* ($n = 47$) across the 42.5 Mb region, and that of *b6*/*b6* throughout the remaining genome. Animals were first subjected to a 10 d macronutrient selection test (see Section 2.4), followed by an extended wash-out period on rodent chow. The macronutrient-rich diets were then reinitiated for 48 h prior to euthanasia and tissue harvest. The 48 h time point for tissue collection was chosen on the basis of temporal variation in genetic linkage, i.e., the absence of genetic linkage for the carbohydrate-rich diet on day 1, and the presence of linkage on days 2 and following [4]. Metabolic signals arising from food ingestion may act to influence the animals' choice beginning on day 2 of diet exposure. Remarkably, the proportion of carbohydrate + protein versus fat + protein diet selected (kcal%) between the 10 d and 2 d macronutrient diet selection tests was highly correlated ($r = 0.86$, $P < 0.01$), emphasizing the stability of the phenotype (preferential carbohydrate intake). The GEO repository contains the data obtained from individuals that displayed the most

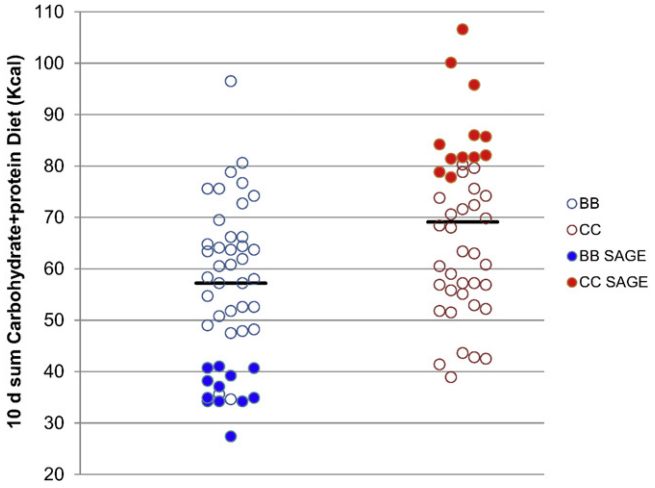

Fig. 1. Scatter plot. Ten day sum of carbohydrate + protein diet (kcal) data for animals of both genotypes including the respective group medians (horizontal black bars). Animals selected their food intake from a choice between two diets: carbohydrate + protein mixture vs. fat + protein mixture, thus the animals' total kcal intake is not represented by this plot. The animals used to generate the SAGE-seq data (filled circles) were chosen from a larger group of animals (open circles), based on intake of the carbohydrate + protein diet. Filled circles represent upper quartile data from carbohydrate-preferring congenic *cast*/*cast* F2 mice (CC, $n = 12$) and lowest quartile data from congenic *b6*/*b6* F2 s (BB, $n = 12$). The B6·CAST-17.1-derived F2 s harbored a non-recombined CAST donor interval (Chr 17:3.19–45.73 Mb) genotype of *cast*/*cast* (red) or *b6*/*b6* (blue), and that of *b6*/*b6* across the remaining genome.

extreme values for self-selected intake of the carbohydrate + protein diet mixture (versus fat + protein diet) in the 10 d test (10 d sum of kcal): *cast/cast* congenic F2 mice in the upper quartile ($n = 12$) and *b6/b6* congenic F2 s in the lowest quartile ($n = 12$) (Fig. 1). Due to a smaller RNA integrity number (RIN), two samples from the 25th quartile for each genotype were replaced with samples associated with the next highest (*cast/cast*) or lowest (*b6/b6*) phenotypic value. Total RNA was isolated from whole hypothalamus, collected from individual animals. RNA sequencing analysis of gene expression was performed on these 24 samples and genes that were differentially expressed between the two genotypes were identified.

2.6. RNA extraction and quality control

Total RNA from whole hypothalamus was extracted and purified using the AllPrep RNA/Protein kit (Qiagen, Valencia, CA) and resuspended in Nanopure water. RNA quality was assessed via nanodrop (Thermo Scientific, Waltham, MA) and Agilent 2100 Bioanalyzer (Santa Clara, CA); all samples showed an RIN of ≥7.0 with the exception of two *cast/cast* samples (6.6, 6.8) and one *b6/b6* sample (6.6).

2.7. cDNA library preparation, SAGE-sequencing & transcriptome analysis

Transcriptome profiling was performed by 3′-expression tag sequencing (SAGE) on an Applied Biosystems (AB) SOLiD 5500XL next generation sequencer. Briefly, sequencing libraries containing 27-bp, 3′ tags for all transcripts within a sample were constructed from hypothalamus using the SOLiD SAGE kit from the manufacturer (Life Technologies), at the PBRC Genomics Core Facility. Each library was then labeled with a unique barcode sequence. Sequence mapping was performed using a modified version of the SOLiD SAGE Analysis Software v1.10 (Life Technologies) and defined analysis parameters. Sequence reads were aligned to mouse RefSeq transcripts (genome build GRCm38/mm10) as the reference. Tag hits, i.e., successfully aligned reads, were normalized or adjusted for coverage according to DESeq by estimating the size factor (median of the ratios of observed counts) for each sample library, and dividing the sample counts by the corresponding size factor [7]. The mapping statistics are summarized in Table 1; only uniquely mapped sequence reads were included in the expression count for each RefSeq gene.

2.8. Differential expression analysis

A principal components analysis (PCA) of gene expression data identified directions or principal components comprising the largest variation in the data [8]. As shown in the two-dimensional PCA biplot (Fig. 3), this analysis revealed two cluster-like patterns of overall gene expression levels characteristic of the congenic *cast/cast* and *b6/b6* F_2 samples. These patterns demonstrate a large influence by genotype on the gene expression profile of hypothalamic cells.

Differential expression analysis of count data was carried out by means of DESeq, an R/Bioconductor package [7]. Genes having both ≥1.5 fold change in expression and $P < 0.01$, were considered to be significantly differentially expressed in the congenic *cast/cast* F_2 compared to congenic *b6/b6* F_2 samples. Two-way hierarchical clustering, via Ward's minimum variance criterion method, was applied to normalized and standardized expression data using the tools in JMP Genomics, Version 10 (SAS Institute Inc., Cary, NC) [5]. The SAGE-seq data were transformed into a scatter plot of the log2 ratio (fold change) versus the mean (Fig. 2). A symmetrical distribution for gene expression was obtained, validating the data quality.

Consistent with the genetic model, the 42.5 Mb Chr 17 *cast/cast* congenic donor interval was enriched for differentially expressed genes. Using filtering criteria of FC ≥1.5 and $P < 0.01$, we identified 86 differentially expressed genes within this interval (Chr 17:3.19–45.73 Mb), while none outside this interval met these criteria. Of the

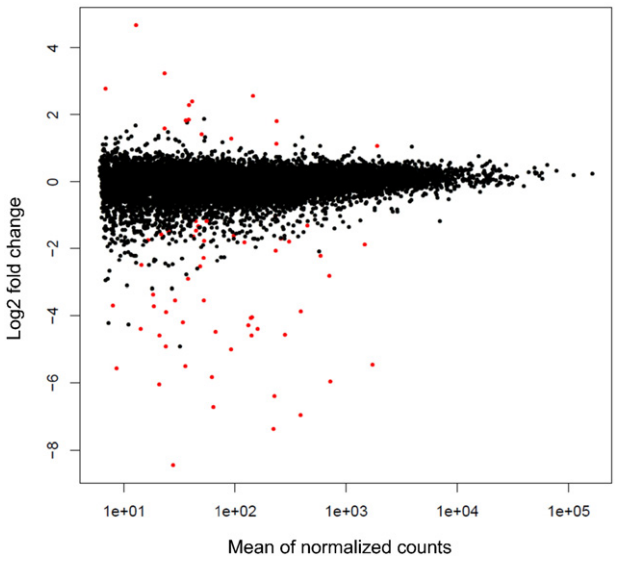

Fig. 2. MA plot. Plot of the log2 fold changes (M), based on a two-group comparison (genotype) over the mean average (A) of normalized counts.

86 differentially expressed genes, 21 exhibited increased expression and 65 were decreased in congenic *cast/cast* F2s compared to congenic *b6/b6* F_2. When we employed a less stringent FC criterion of ≥1.2, our analysis revealed 1–3 differentially expressed genes on each of eleven other autosomes, for a total of 18 DE genes outside the boundaries of the *cast/cast* congenic donor region, defined as Chr 17:3.19–45.73 Mb by SNP genotyping [5]. Differentially expressed genes located outside this 42.5 Mb congenic region are postulated to be under the control of *trans*-acting genetic factors.

3. Discussion

Here we describe our methods of hypothalamic transcriptome analysis to generate a list of differentially expressed genes in carbohydrate-

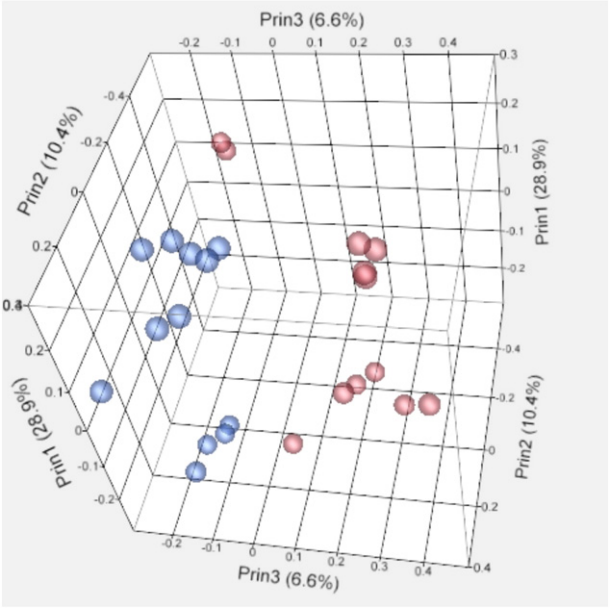

Fig. 3. Scatter plot of PCA for 24 transcript profiles. A principal components analysis (PCA) of gene expression data identified principal components comprising the largest variation in transcript profiles for B6·CAST-17.1-derived F2 samples possessing a non-recombined congenic interval (Chr 17:3.19–45.73 Mb) genotype of *cast/cast* (red, $n = 12$)) or *b6/b6* (blue, $n = 12$), and that of *b6/b6* across the remaining genome.

preferring B6·CAST-17.1 *cast*/*cast* congenic F2 mice compared to fat-preferring B6-CAST-17.1 *b6*/*b6* congenic F2, in which both genotypes possess an otherwise *b6*/*b6* genome. The Chr 17 CAST congenic segment encompasses the fine-mapped *Mnic* QTL (Chr 17:26.08–45.12 Mb, 95% confidence interval) that is associated with the preferential carbohydrate intake of *cast*/*cast* animals in a macronutrient diet choice paradigm [5]. The hypothalamic transcriptome represented by this fine-mapped 19.0 Mb region contains 55 differentially expressed genes, of which nearly 50% were classified as metabolism genes by gene ontology analysis [5]. Seven of these genes are associated with lipids and/or fatty acids and showed decreased expression in fat-preferring *b6*/*b6* congenic F2 (*Decr2*, *Ppard*, *Agpat1*, *Tnxb*, *Neu1*, *Pla2g7*, *Cyp39a1*). By contrast, two genes with increased expression in carbohydrate-preferring *cast*/*cast* congenic F2 mice have relevance to carbohydrate metabolism (*Glo1*, *Neu1*). Notably, three of these same genes (*Agpat1*, *Pla2g7*, *Glo1*) were found to exhibit significantly differential expression, in the same direction and with similar magnitude, in the hypothalamus of other congenic lines containing a CAST Chr 17 segment on a B6 background [2,4].

Funding

This research was supported by the National Institutes of Health (NIH) (DK-53113 to BKR). The work utilized the Genomics Core Facilities (GCF) which is partially supported by Center of Biomedical Research Excellence (COBRE) NIH 8 P20 GM103528 and Nutrition Obesity Research Center (NORC) NIH P30DK072476. We thank Dr. Michael Salbaum, Director of the GCF at Pennington Biomedical for helpful advice on the SAGE experiment.

References

[1] H.R. Berthoud, H. Munzberg, B.K. Richards, C.D. Morrison, Neural and metabolic regulation of macronutrient intake and selection. Proc. Nutr. Soc. 71 (2012) 390–400.
[2] K.G. Kumar, L.M. DiCarlo, J. Volaufova, A.R. Zuberi, B.K. Richards, Increased physical activity cosegregates with higher intake of carbohydrate and total calories in a subcongenic mouse strain. Mamm. Genome 21 (2010) 52–63.
[3] K.G. Kumar, A.C. Poole, B. York, J. Volaufova, A. Zuberi, B.K. Richards, Quantitative trait loci for carbohydrate and total energy intake on mouse chromosome 17: congenic strain confirmation and candidate gene analyses (Glo1, Glp1r). Am. J. Phys. Regul. Integr. Comp. Phys. 292 (2007) R207–R216.
[4] K.G. Kumar, B.K. Smith Richards, Transcriptional profiling of chromosome 17 quantitative trait loci for carbohydrate and total calorie intake in a mouse congenic strain reveals candidate genes and pathways. J Nutrigenet Nutrigenomics 1 (2008) 155–171.
[5] R. Gularte-Merida, L.M. DiCarlo, G. Robertson, J. Simon, W.D. Johnson, C. Kappen, J.F. Medrano, B.K. Richards, High-resolution mapping of a genetic locus regulating preferential carbohydrate intake, total kilocalories, and food volume on mouse chromosome 17. PLoS One 9 (2014), e110424.
[6] B.K. Smith Richards, B.N. Belton, A.C. Poole, J.J. Mancuso, G.A. Churchill, R. Li, J. Volaufova, A. Zuberi, B. York, QTL analysis of self-selected macronutrient diet intake: fat, carbohydrate, and total kilocalories. Physiol. Genomics 11 (2002) 205–217.
[7] S. Anders, W. Huber, Differential expression analysis for sequence count data. Genome Biol. 11 (2010) R106.
[8] M. Ringner, What is principal component analysis? Nat. Biotechnol. 26 (2008) 303–304.