Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Musunuru K, Pirruccello JP, Do R, et al. Exome sequencing, *ANGPTL3* mutations, and familial combined hypolipidemia. N Engl J Med 2010;363:2220-7.

SUPPLEMENTARY APPENDIX

EXOME SEQUENCING

Exome sequencing was performed at the Broad Institute. DNA oligonucleotides, corresponding to 170 bp of target sequence flanked by 15 bp of universal primer sequence, were synthesized in parallel on an Agilent microarray, then cleaved from the array. The oligonucleotides were PCR amplified, then transcribed *in vitro* in the presence of biotinylated UTP to generate single-stranded RNA "bait." Genomic DNA was sheared, ligated to Illumina sequencing adapters, and selected for lengths between 200-350 bp. This "pond" of DNA was hybridized with an excess of bait in solution. The "catch" was pulled down by magnetic beads coated with streptavidin, then eluted.[1]

Each sample was sequenced on three lanes of an Illumina GA-II sequencer, using 76 bp paired-end reads.

EXOME SEQUENCE ANALYSIS

A pipeline based on the Picard suite (http://picard.sourceforge.net/) was used to process the unaligned Illumina reads and to emit aligned reads in the Sequence Alignment/Map (SAM) specification. Unaligned reads were aligned to the human genome (HG18) using Maq (http://maq.sourceforge.net/) in a two-pass process: first, with low-stringency for calibration of base quality scores and, second, with high-stringency for alignment of all high-quality reads. Quality scores for each base were recalibrated using the Genome Analysis Toolkit (GATK).[2] Reads not corresponding to the 28,646,006 targeted bases of the exome were then discarded.

Sixty exomes prepared from non-tumor tissue DNA of patients diagnosed with either multiple myeloma (N=20) or glioblastoma multiforme (N=40) and processed at the Broad Institute using the same solution hybrid capture, sequencing, and analysis pipeline were used as controls.

The SAM files of 66 exomes (including those of the two individuals discussed in this paper, as well as 4 unrelated individuals with low LDL cholesterol and the 60 controls) were then subjected to an analytical pipeline for high-confidence variant calling and annotation. All 66 exomes were shown to GATK's UnifiedGenotyper in multisample mode. UnifiedGenotyper performs variant calling under a Bayesian framework that identifies sites likely to be variant, and then assigns the most probable genotype (and Phred-scaled confidence score for that genotype) for each individual at those variant sites given the error model for the sequencing method being used, the evidence accumulated in the reads, and the likely heterozygosity at the locus.[2] The GATK's VariantFiltrationWalker then flags for downstream analysis those variants that met the following quality control requirements: (a) number of aligned reads at the locus (depth of coverage) greater than 7; (b) ratio of the variant confidence score to the depth of coverage greater than 5; and (c) nonreference allele present in at least 35% of reads at the locus in individuals for whom the variant was called.

The remaining high-confidence variants were annotated for functional consequence based on RefSeq transcripts (http://www.ncbi.nlm.nih.gov/refseq/). The variants were annotated for novelty based on non-overlap with variants present in dbSNP 130

(http://www.ncbi.nlm.nih.gov/projects/SNP/), 1000 Genomes (see below), or the 60 control exomes (data not shown).

1000 Genomes single nucleotide variants (SNVs) were obtained from a merger of two datasets. The first set was generated by Richard Durbin and colleagues, using phased haplotypes (ftp://ftp.sanger.ac.uk/pub/1000genomes/REL-0908/LowCov/) of 56 CEU individuals (NA12891, NA12878, NA10847, NA10851, NA12004, NA12414 and NA12717 were removed) from the August 2009 data release of the 1000 Genomes Project. The second set was generated by Yun Li, Goncalo Abecasis, and colleagues using phased haplotypes (ftp://ftp.sanger.ac.uk/pub/1000genomes/REL-0908/LowCov/) of 61 CEU individuals (NA12891 and NA12878 were removed) from the same data release. Singletons were removed from both datasets.

VALIDATION USING SANGER SEQUENCING

In all 38 family members for whom DNA was available, exon 1 of *ANGPTL3* was amplified using the following primers: 5′-ACCTACCAACCTTACCTTTTCTGGGCA-3′ (forward strand) and 5′-TCTGGGTGTTCTGGAGTTTCAGGT-3′ (reverse strand). This encompasses the region between position 62,835,710 and position 62,836,305 on chromosome 1. The product was then subjected to Sanger sequencing to determine the presence or absence of the S17X and E129X mutations.

LINKAGE ANALYSES

A total of 378 simple sequence repeat (SSR) markers were genotyped in the 38 individuals from

the FHBL family by the Mammalian Genotyping Service in Marshfield Clinic. SOLAR[3] was used to generate variance component models with quantitative measurements of LDL cholesterol or HDL cholesterol, square-root transformed to achieve normality. Sex, age, and body-mass index were considered as covariates. Two-point linkage analysis was performed on the SSR markers, and the highest LOD score was seen between markers GATA72H07 and GATA109 (on chromosome 1p32.3) for LDL cholesterol and HDL cholesterol (**Supplementary Appendix Table 2**). After identifying the *ANGPTL3* nonsense variants, two separate linkage analyses were performed: first, the nonsense variants were included as a covariate, coded by the number of variants (0, 1, or 2); second, linkage analysis was performed using the *ANGPTL3* nonsense variants in addition to the SSR markers (**Supplementary Appendix Table 2**).

The peak linkage interval (between GATA72H07 and GATA109) spans ~34 million bases and harbors 165 refGene-annotated genes including *ANGPTL3*. Prior to exome sequencing, we had performed targeted sequencing of candidate genes with plausible links to LDL cholesterol metabolism present in the linkage interval—*PCSK9*, *SCP2*, *CPT2*, *OSBPL9*, *LRP8*, and *CYP2J2*—and found no novel variants in any of these genes. With exome sequencing, we found that besides *ANGPTL3*, no genes in the interval harbored novel variants in both alleles that were shared by the two probands. Only four other genes in the interval harbored novel variants, but they were present only on one allele in each of the most severely affected individuals in the family (**Supplementary Appendix Table 3**).

When we performed the linkage analyses using the number of *ANGPTL3* nonsense variants as a covariate, the LOD scores at the peak microsatellite markers became non-significant

(**Supplementary Appendix Table 2**). When we modeled the *ANGPTL3* nonsense variants as a marker and added it to the linkage analyses, we found it to have the highest LOD score for LDL cholesterol (4.63) (**Supplementary Appendix Table 2**).

REPLICATION IN A POPULATION-BASED COHORT

Sequencing *ANGPTL3* in 3,551 individuals (7,102 chromosomes) in the Dallas Heart Study[14] failed to reveal any nonsense mutations. However, 12 individuals were heterozygous for frameshift mutations in *ANGPTL3*: FsY83 (n = 1), FsS122 (n = 2), FsN147 (n = 2), FsQ192 (n = 5), FsK389 (n = 1), and FsK445 (n = 1). Carriers of frameshift mutations (each leading to premature truncation of the protein and, thus, functionally similar to nonsense mutations) had significantly lower plasma levels of LDL cholesterol than did non-carriers (median 77.5 mg/dl in carriers vs. median 104 mg/dl in non-carriers; P = 0.03) (**Supplementary Appendix Table 4**). Mutation carriers also had lower triglyceride levels, albeit of borderline significance (median 72 mg/dl in carriers vs. median 96 mg/dl in non-carriers; P = 0.08); no difference was seen for HDL cholesterol (P = 0.53).

LIPOPROTEIN METABOLIC STUDIES

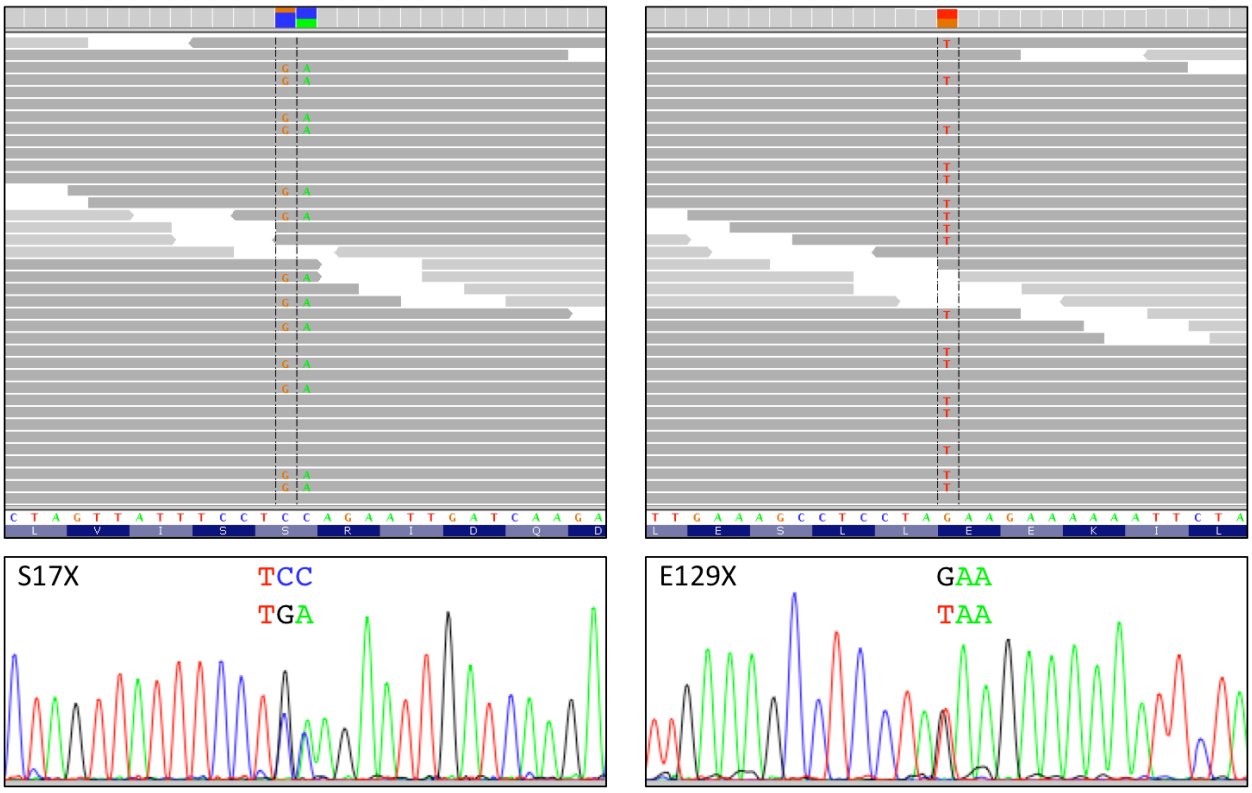
Previously published physiologic studies of selected members of our family[4] suggested that affected individuals have both decreased production rates of VLDL apolipoprotein B and increased fractional catabolic rates for LDL apolipoprotein B. We are now able to categorize these individuals by carrier status for *ANGPTL3* mutations. With this categorization, we found a gene-dosage effect for VLDL production rates (decreased rates with additional mutations, P =

0.001) and LDL fractional catabolic rates (increased rates with additional mutations, $P = 0.005$) (**Supplementary Appendix Table 5**).

SUPPLEMENTARY REFERENCES

1. Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009;27:182-9.
2. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, in the press.
3. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 1998;62:1198-211.
4. Elias N, Patterson BW, Schonfeld G. In vivo metabolism of ApoB, ApoA-I, and VLDL triglycerides in a form of hypobetalipoproteinemia not linked to the ApoB gene. Arterioscler Thromb Vasc Biol 2000;20:1309-15.
5. Yuan B, Neuman R, Duan SH, et al. Linkage of a gene for familial hypobetalipoproteinemia to chromosome 3p21.1-22. Am J Hum Genet 2000;66:1699-704.

Supplementary Appendix Figure 1.



Integrated Genome Viewer (IGV) views of exome sequencing reads for each *ANGPTL3* variant in individual II.5 aligned with Sanger sequencing traces. Each horizontal gray bar represents a single 76 base pair sequence read. Individual II.5 is a compound heterozygote for both nonsense mutations—S17X and E129X. S17X is the result of a two base substitution (TCC→TGA) that changes serine at position 17 to a stop codon. E129X is the result of a single base substitution (GAA→TAA) that changes glutamic acid at position 129 to a stop codon.

Supplementary Appendix Table 1. Phenotypic characteristics of members of the family from the current study.

Individual	Identifier*	Sex	Age	BMI	Glucose	LDL-C	HDL-C	TG	TC	ApoAI	ApoB	S17X†	E129X†
I-1	201	F	68	32.35	93	134	33.3	174.3	202	104	103	0	0
I-2	202	M	69	28.19	90	156.3	37	125.7	218.7	113	111.5	0	0
I-3	203	M	63	16.54	99	65.7	33.3	83	115.7	102	62.5	1	0
I-4	204	F	61	35.22	91	102.5	39	316	204.5	152	118	0	0
I-5	205	M	74	27.95	90	95.7	51	114.3	169.7	145	78.5	1	0
I-6	206	M	60	24.01	99	105.7	30.3	148	165.7	108	86.5	0	0
I-7	208	F	64	30.73	91	73.2	44.2	51	126	128	59.8	1	0
I-8	207	M	67	32.35	81	98.8	35	97.8	145.4	121	81	0	1
II-1	301	M	40	32.35	—	155	39	378	270	131	139	0	0
II-2	306	M	43	27.38	83	34.7	16.3	24.3	56	51.5	41.5	1	1
II-3	307	F	35	21.97	—	112.3	38.7	63.7	164.3	110	83.5	0	0
II-4	302	M	23	28.51	101	30.8	15.8	22.3	50.5	51.5	33.7	1	1
II-5	303	M	28	25.14	102	29	22	17	54	65	27	1	1
II-6	305	F	44	26.68	85	36.8	17.8	19.1	57.4	47.6	37.9	1	1
II-7	304	M	49	27.18	—	214.3	40.3	343.7	313	158	194	0	0
II-8	311	F	36	25.85	—	87	35.5	66.7	136.7	124	70.5	1	0
II-10	309	F	42	22.52	88	68.2	48.2	44.8	127	105.6	59.3	0	1
II-11	308	M	44	29.15	—	116.3	36.3	90.7	171.7	107	92	0	0
II-12	312	M	39	32	—	121.8	49.4	53.7	187	125.5	95	0	0
II-13	313	F	36	23.57	—	110.5	49	92	177.5	127	95	0	0
II-14	318	M	40	32.84	—	147.6	37.6	224.2	227.3	130.3	126	0	0
II-15	314	M	36	31.24	108	88.5	36.8	87.6	133.2	116	72	1	0
II-16	315	F	37	24.75	81	85	89	80	190	186	71	0	0
II-17	317	F	34	22.36	79	73.9	57.4	42.1	138.3	133.8	56	0	0
III-1	403	M	18	24.84	112	28.3	42.3	42.7	79.7	94	26	1	0
III-2	404	F	14	15.61	121	40	53	17	99	115	31	0	1
III-3	405	M	10	16.3	114	54.3	51.3	31.7	111	121	44.5	0	1
III-4	401	F	18	20.72	103	71.5	49	55.5	129.8	134.5	60.5	1	0
III-5	402	F	16	23.42	86	102.7	34.7	106.8	154.8	106.5	85.4	0	1
III-6	407	M	14	24.8	—	61.2	37	144	123	139	69.3	0	0
III-7	408	F	11	31.53	—	65.3	37.8	145.5	133	124	67	0	0
III-8	406	M	18	17.26	90	61.4	40.4	29.8	107.2	104.3	49.3	0	1
III-9	409	F	20	25.74	—	84.5	45.3	56.3	140.5	120	65.3	0	0
III-10	410	M	13	23.45	—	97.3	45.7	77.7	158.3	125.5	81.5	0	0
III-11	411	F	17	22.3	—	86.7	47.7	43.7	142.7	117	76	0	0
III-12	412	F	10	—	—	91.5	46	45	146.5	119	68.5	0	0
III-13	413	M	11	15.71	—	75	74	31	155	166	67	0	0
III-14	414	M	9	19.14	—	92	62	49.8	164.2	153	71	0	0

BMI = body-mass index in kg/m^2 units; glucose in mg/dL units; LDL-C = low-density lipoprotein cholesterol in mg/dL units; HDL-C = high-density lipoprotein cholesterol in mg/dL units; TG = triglycerides in mg/dL units; TC = total cholesterol in mg/dL units; ApoAI = apolipoprotein A-I in mg/dL units; ApoB = apolipoprotein B in mg/dL units.

* Identifier used in pedigree in ref. 5 (ref. 6 in the main manuscript). † 0 indicates no mutation was present; 1 indicates mutation was present.

Supplementary Appendix Table 2. Two-point linkage analyses on HDL cholesterol and LDL cholesterol at markers on chromosome 1, with or without the *ANGPTL3* mutations as markers.

			LOD Scores					
			HDL cholesterol			LDL cholesterol		
Chr	Position	Marker name	Without mutations	Mutations as covariates	Mutations as markers	Without mutations	Mutations as covariates	Mutations as markers
1	3,484,862-3,685,185	280we5	0	0.003	0	-0.0001	0	0
1	8,017,839-8,218,246	GGAA3A07	0	0.3809	0	0	0.4581	0
1	13,684,108-13,884,418	GATA27E01	0	0.0029	0	0.1149	0.5128	0.1354
1	17,683,598-17,883,916	GATA29A05	0.0076	0.2948	0.0076	0.2695	0	0.266
1	19,166,961-19,367,289	GATA124F08	0.5238	0.003	0.5238	0.6451	0.1794	0.6073
1	19,166,961-19,367,289	GGAT2A07	0.1195	0.1249	0.1195	-0.0001	0	0
1	30,210,927-30,411,242	ATA20F08	0	0.0028	0	0	0	0
1	41,626,467-41,826,745	GATA129H04	1.0544	0.0029	1.0544	1.3607	0	1.311
1	48,181,221-48,381,554	GATA72H07	0.91	0.0023	0.91	2.2725	0.7045	2.2235
1	**60,571,863-60,772,183**	**GATA165C03**	**2.7355**	**0.331**	**2.7355**	**3.0653**	**0**	**3.0164**
1	**63,063,187-63,071,180**	***ANGPTL3* mutations**	**—**	**—**	**2.6516**	**—**	**—**	**4.6369**
1	**74,129,381-74,329,742**	**GATA61A06**	**1.305**	**0.003**	**1.305**	**3.8889**	**1.0837**	**3.9276**
1	81,898,598-82,098,757	GATA109	0.4847	0.1141	0.4847	0.4601	0.4408	0.4589
1	82,795,173-82,995,466	GATA6A05	0.5325	0.1421	0.5325	0.2716	0	0.2795
1	92,114,212-92,314,568	ATA2E04	0	0.003	0	0	0	-0.0004
1	105,560,652-105,760,779	ATA29D04	0	0.0028	0	-0.0006	0	-0.0005
1	107,554,572-107,754,850	GATA176G01	0	0.003	0	-0.0006	0	0
1	119,578,203-119,778,621	GATA12A07	0	0.0027	0	0	0	0.0017
1	157,832,671-158,032,993	GATA43A04	0.0275	0.3795	0.0275	0	0	0
1	162,261,764-162,462,074	GGAA5F09	0	0.003	0	-0.0006	0	-0.0001
1	163,459,700-163,660,041	GGAA22G10	0	0.0385	0	0	0	0
1	174,161,084-174,361,415	ATA4E02	0.0297	0.0029	0.0297	0.3786	0	0.3533
1	187,450,180-187,650,535	GATA7C01	0	0.0028	0	0	0	0
1	198,511,279-198,711,576	GATA48B01	0	0.003	0	0.1502	0.8219	0.1395
1	203,431,706-203,632,058	GGAA23C07	0.015	0.0022	0.015	0.7704	1.3231	3.0687
1	215,095,283-215,295,598	GATA87F04	0.6507	0.003	0.6507	1.2986	0	1.3058
1	219,543,613-219,743,932	GATA4H09	0.4339	0.003	0.4339	0.8998	0	0.9045
1	231,850,486-232,050,793	ATA29C07	0.4684	0.0028	0.4684	0.1846	0	0.1938
1	235,793,755-235,994,050	203yg9	0.9128	0.0022	0.9128	0.5877	0	0.8299
1	241,655,914-241,856,231	GATA4A09	1.5962	0.0029	1.5962	0.4418	0	0.5262
1	243,965,857-244,166,112	GATA50F11	0.0801	0.003	0.0801	0	0.0378	-0.0001

Supplementary Appendix Table 3. Novel nonsynonymous variants under the chromosome 1 linkage peak.

Chr	Pos	Gene	Ref. Codon	Ref. Amino Acid	Codon Number	Alt. Codon	Alt. Amino Acid
1	55053290	*C1orf177*	ATG	Met	347	ACG	Thr
1	62689091	*USP1*	GTG	Val	737	ATG	Met
1	62835875/ 62835876	*ANGPTL3*	TCC	Ser	17	TGA	Opl
1	62836210	*ANGPTL3*	GAA	Glu	129	TAA	Ocr
1	76136268	*MSH4*	CAA	Gln	815	CGA	Arg
1	78873759	*IFI44L*	TAT	Tyr	291	TAA	Ocr

Supplementary Appendix Table 4. Phenotypic characteristics of individuals from the Dallas Heart Study with frameshift mutations in *ANGPTL3*.

Sex	Ethnicity*	Age	BMI	LDL-C	HDL-C	TG	TC	Glucose	Frameshift location	Stop codon location	Frameshift size (nucleotides)	Frameshift type
F	AA	47	32.3	56	48	77	120	88	Y83	X99	1	Deletion
F	AA	32	44.0	150	54	42	213	98	S122	X128	4	Deletion
F	C	35	28.1	50	46	51	106	92	S122	X128	4	Deletion
F	AA	42	—	77	47	47	134	83	N147	X148	4	Deletion
F	C	56	53.7	78	34	162	144	137	N147	X148	4	Deletion
M	AA	48	23.6	77	29	143	134	115	Q192	X196	1	Deletion
M	AA	34	21.9	43	65	50	118	102	Q192	X196	1	Deletion
F	AA	59	26.0	113	68	73	196	95	Q192	X196	1	Deletion
M	AA	42	25.8	99	61	71	174	106	Q192	X196	1	Deletion
M	AA	40	29.8	106	65	74	186	80	Q192	X196	1	Deletion
M	C	42	25.0	84	50	82	150	86	K389	X413	1	Deletion
M	AA	50	34.6	76	68	64	157	83	K445	X459	1	Insertion
Average		44	31.4	84	53	78	153	97				

BMI = body-mass index in kg/m^2 units; LDL-C = low-density lipoprotein cholesterol in mg/dL units; HDL-C = high-density lipoprotein cholesterol in mg/dL units; TG = triglycerides in mg/dL units; TC = total cholesterol in mg/dL units; glucose in mg/dL units.

* AA = African-American; C = Caucasian.

Supplementary Appendix Table 5. Lipoprotein metabolic studies in selected individuals.

Individual	S17X†	E129X†	VLDL apoB PS, mg/kg	VLDL apoB FCR, pools/h	VLDL apoB PR, $mg \times kg^{-1} \times d^{-1}$	IDL apoB FCR, pools/h	LDL apoB PS, mg/kg	LDL apoB FCR, pools/h	LDL apoB PR, $mg \times kg^{-1} \times d^{-1}$	VLDL TG PS, µmol/kg	VLDL TG FCR, pools/h	VLDL TG PR, $\mu mol \times kg^{-1} \times h^{-1}$
II-4	1	1	0.70	0.51	8.6	1.36	8.0	0.043	8.2	4.1	2.14	8.8
II-6	1	1	0.28	1.26	8.5	2.75	5.5	0.061	8.1	1.5	3.41	5.1
II-10	0	1	0.13	3.5	10.9	4.33	6.5	0.055	9.1	3.1	4.68	14.4
II-15	1	0	1.76	0.48	20.5	0.40	19.3	0.039	17.9	18.8	1.72	32.3
III-1	1	0	0.91	0.81	17.7	0.37	5.5	0.041	6.0	9.1	0.51	4.6
II-12	0	0	0.54	2.33	30.2	0.41	16.9	0.028	11.4	17.8	1.70	30.3
III-11	0	0	1.27	0.69	21.0	0.44	15.3	0.024	9.0	14.7	0.90	13.2
N1*	0	0	1.67	0.64	25.6	0.39	25.2	0.025	15.1	15.3	0.69	10.6
N2*	0	0	0.50	2.57	30.8	0.54	13.2	0.032	10.3	10.6	1.08	11.4
Mean ± SD	2 mutations		0.5 ± 0.3	0.9 ± 0.5	8.6 ± 0.1	2.1 ± 1.0	6.8 ± 1.8	0.052 ± 0.013	8.2 ± 0.1	2.8 ± 1.8	2.8 ± 0.9	7.0 ± 2.6
	1 mutation		0.9 ± 0.8	1.6 ± 1.7	16.4 ± 4.9	1.7 ± 2.3	10.4 ± 7.7	0.045 ± 0.009	11.0 ± 6.2	10.3 ± 7.9	2.3 ± 2.1	17.1 ± 14.0
	0 mutations		1.0 ± 0.6	1.6 ± 1.0	26.9 ± 4.6	0.5 ± 0.07	17.7 ± 5.3	0.027 ± 0.004	11.5 ± 2.6	14.6 ± 3.0	1.1 ± 0.4	16.4 ± 9.3
P-value, additive model of genetic dosage			0.39	0.57	0.001	0.162	0.05	0.005	0.38	0.03	0.14	0.38

PS indicates pool size; PR, production rate; FCR, fractional catabolic rate.

* N1 was the spouse of II-6 and was not included in the pedigree; N2 was a son of II-12 and was not included in the pedigree.

† 0 indicates no mutation was present; 1 indicates mutation was present.