SUPPLEMENTARY TABLE 1. PHENOTYPIC SUBGROUP ANALYSIS OF CONFIRMED LOCI

						Replication #1				Replication #2				
						Ileal CD: 530 trios				Ileal CD: 353 Control: 207				
MARKER #	Chr	Position	SNP	Locus	GWA p-value	T:U	OR	Chi2	P	Case MAF	Cont MAF	OR	Chi2	P
18	2	233965368	rs2241880	ATG16L1 (T197A)	1.063E-08	219:305	0.718	14.11	0.000086	0.353	0.478	0.595	17.054	0.000018
41	4	41594058	rs16853571	Phox2	2.396E-07	39:75	0.52	11.37	0.0003735	0.058	0.047	1.241	0.572	0.45
55	10	64140681	rs224136	intergenic	0.000001852	93:149	0.6242	12.96	0.0001592	0.14	0.23	0.544	14.585	0.000134
59	13	74229094	rs11617463	intergenic	0.00001617	59:80	0.7375	3.173	0.03744					
68	16	83696674	rs8050910	FAM92B	0.000009546	221:268	0.8246	4.517	0.016775	0.4	0.43	0.884	0.884	0.347
75	22	35583003	rs4821544	NCF4	0.00001753	266:221	1.204	4.158	0.02072	0.374	0.339	1.164	1.331	0.249

						CD: 619 trios				CD: 625 Control: 207				
MARKER #	Chr	Position	SNP	Locus	GWA p-value	T:U	OR	Chi2	P	Case_MAF	Cont_MAF	OR	Chi2	P
18	2	233965368	rs2241880	ATG16L1 (T197A)	1.063E-08	263:350	0.7514	12.35	0.0002208	0.373	0.478	0.648	14.424	0.0001
41	4	41594058	rs16853571	Phox2	2.396E-07	57:78	0.7308	3.267	0.03535	0.054	0.047	1.178	0.379	0.538
55	10	64140681	rs224136	intergenic	0.000001852	119:172	0.6919	9.653	0.0009455	0.159	0.23	0.629	10.887	0.001
59	13	74229094	rs11617463	intergenic	0.00001617	76:93	0.8172	1.71	0.0955					
68	16	83696674	rs8050910	FAM92B	0.000009546	259:301	0.8605	3.15	0.037965	0.401	0.423	0.887	1	0.317
75	22	35583003	rs4821544	NCF4	0.00001753	298:274	1.088	1.007	0.1578	0.374	0.339	1.164	1.595	0.207

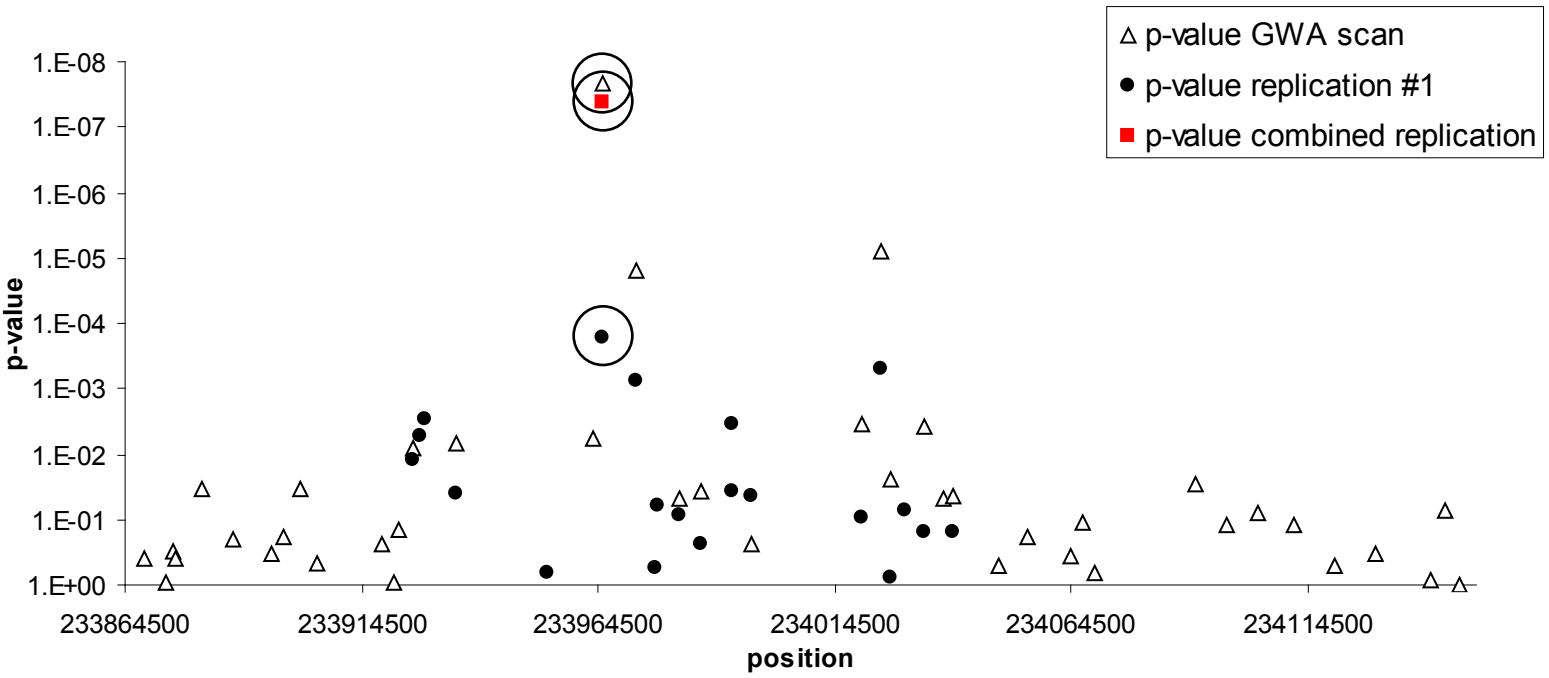
						UC: 249 trios				UC: 353 Control: 207				
MARKER #	Chr	Position	SNP	Locus	GWA p-value	T:U	OR	Chi2	P	Case MAF	Cont MAF	OR	Chi2	P
18	2	233965368	rs2241880	ATG16L1 (T197A)	1.063E-08	109:125	0.872	1.094	0.1478	0.415	0.478	0.773	4.268	0.039
41	4	41594058	rs16853571	Phox2	2.396E-07	34:22	1.545	2.571	0.0544	0.058	0.047	1.262	0.672	0.413
55	10	64140681	rs224136	intergenic	0.000001852	62:70	0.8857	0.4848	0.2431	0.144	0.23	0.563	13.178	0.0003
59	13	74229094	rs11617463	intergenic	0.00001617	41:36	1.139	0.3247	0.2844					
68	16	83696674	rs8050910	FAM92B	0.000009546	129:120	1.075	0.3253	0.2842	0.422	0.423	0.969	0.06	0.807
75	22	35583003	rs4821544	NCF4	0.00001753	97:121	0.8017	2.642	0.05205	0.353	0.339	1.063	0.218	0.641

						IBD: 961 trios				IBD: 978 Control: 207				
MARKER #	Chr	Position	SNP	Locus	GWA p-value	T:U	OR	Chi2	P	Case_MAF	Cont_MAF	OR	Chi2	P
18	2	233965368	rs2241880	ATG16L1 (T197A)	1.063E-08	412:522	0.7893	12.96	0.00015955	0.388	0.478	0.691	11.581	0.0007
41	4	41594058	rs16853571	Phox2	2.396E-07	97:117	0.8291	1.869	0.0858	0.056	0.047	1.208	0.553	0.457
55	10	64140681	rs224136	intergenic	0.000001852	204:274	0.7445	10.25	0.000683	0.153	0.23	0.605	14.377	0.0001
59	13	74229094	rs11617463	intergenic	0.00001617	125:142	0.8803	1.082	0.1491					
68	16	83696674	rs8050910	FAM92B	0.000009546	410:473	0.8668	4.495	0.017	0.408	0.423	0.915	0.592	0.442
75	22	35583003	rs4821544	NCF4	0.00001753	437:431	1.014	0.04147	0.4193	0.366	0.339	1.127	1.079	0.299

Supplementary Table 2. Primer sequences for real-time RT-PCR assays for RNA quantitation experiments

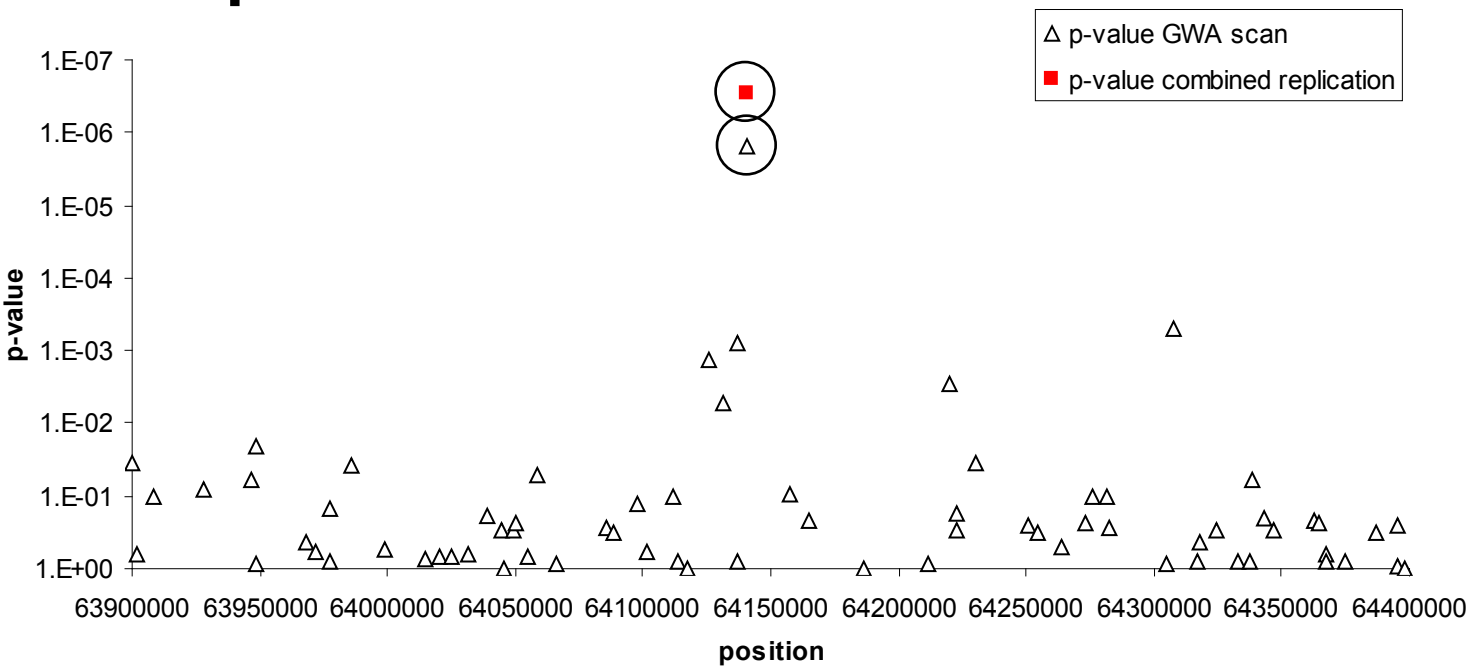
GENE	PRIMER	SEQUENCE
ATG5	Forward	5'-TTG ACG TTG GTA ACT GAC AAA GT-3'
ATG5	Reverse	5'-TGT GAT GTT CCA AGG AAG AGC-3',
ATG7	Forward	5'-GAT CCG GGG ATT TCT TTC ACG-3'
ATG7	Reverse	5'-CAG CAA TGT AAG ACC AGT CAA GT-3'
ATG16L1	Forward	5'-TGA TGG CAC ATG GAA TGA CAA-3'
ATG16L1	Reverse	5'-GAG TCG CTT AGT GGC TGC TC-3'
GAPDH	Forward	5'-GGA GCC AAA CGG GTC ATC ATC TC-3'
GAPDH	Reverse	5'-GAG GGG CCA TCC ACA GTC TTC T-3'

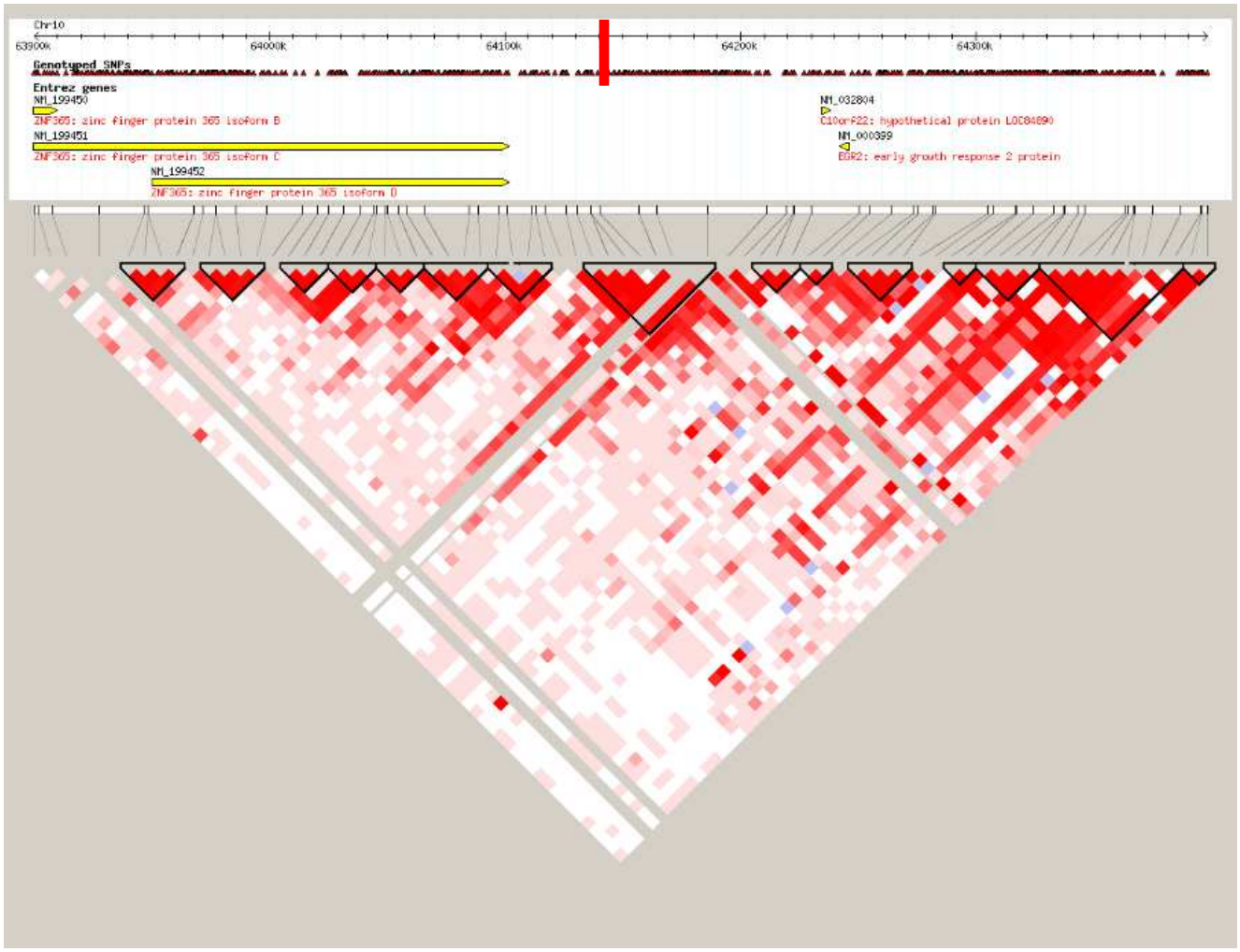
A. ATG16L1


p-value GWA scan
p-value replication #1
p-value combined replication
1.E-08
1.E-07
1.E-06
1.E-05
1.E-04
1.E-03
1.E-02
1.E-01
1.E+00
p-value
233864500
233914500
233964500
234014500
234064500
234114500
position

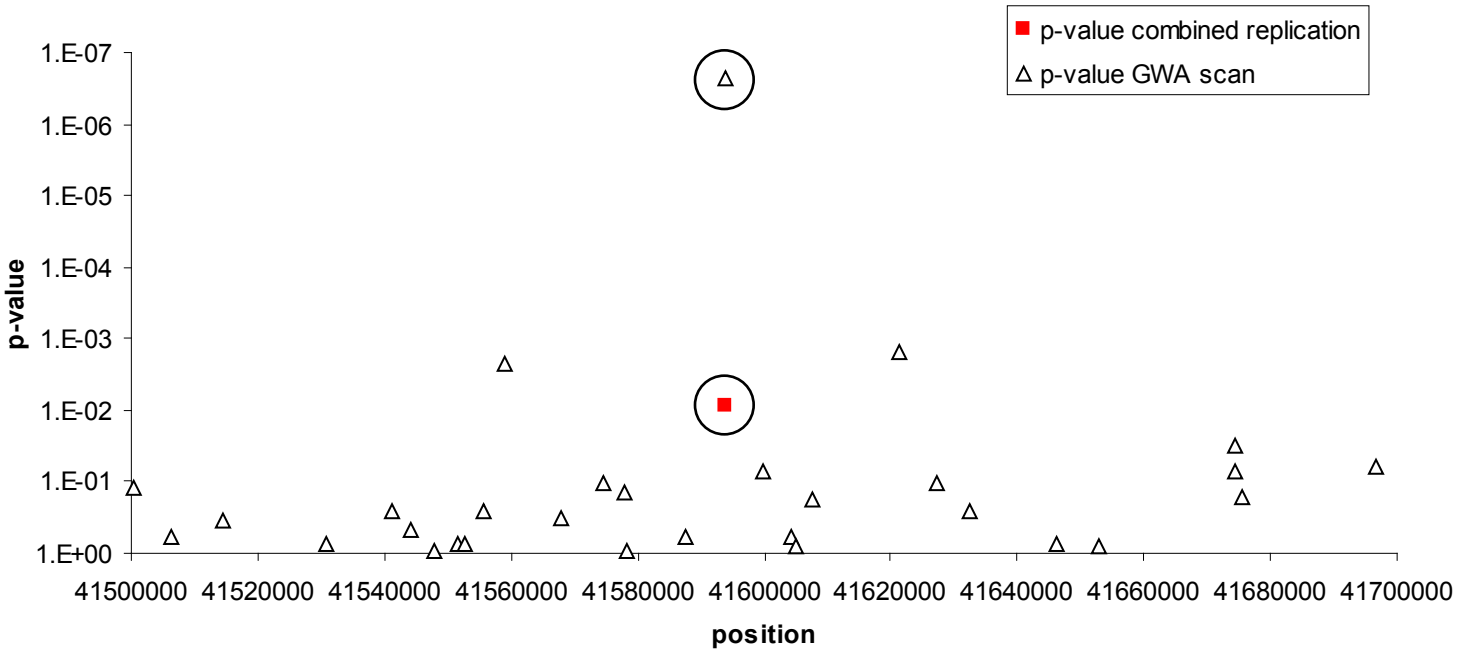

Chr2
233900k
234000k
234100k
Genotyped SNPs
Entrez genes
NM_001017915
INPP5D: SH2 containing inositol phosphatase isoform a
NM_005541
INPP5D: SH2 containing inositol phosphatase isoform b
NM_017974
ATG16L1: APG16 autophagy 16-like isoform 2
NM_030803
ATG16L1: APG16 autophagy 16-like isoform 1
NM_198890
ATG16L1: APG16 autophagy 16-like isoform 3
NM_000541
SAG: S-arrestin
NM_152879
DGKD: diacylglycerol kinase
NM_003648
DGKD: diacylglycerol kinase

B. 10q21.1


p-value GWA scan
p-value combined replication
1.E-07
1.E-06
1.E-05
1.E-04
1.E-03
1.E-02
1.E-01
1.E+00
p-value
63900000
63950000
64000000
64050000
64100000
64150000
64200000
64250000
64300000
64350000
64400000
position

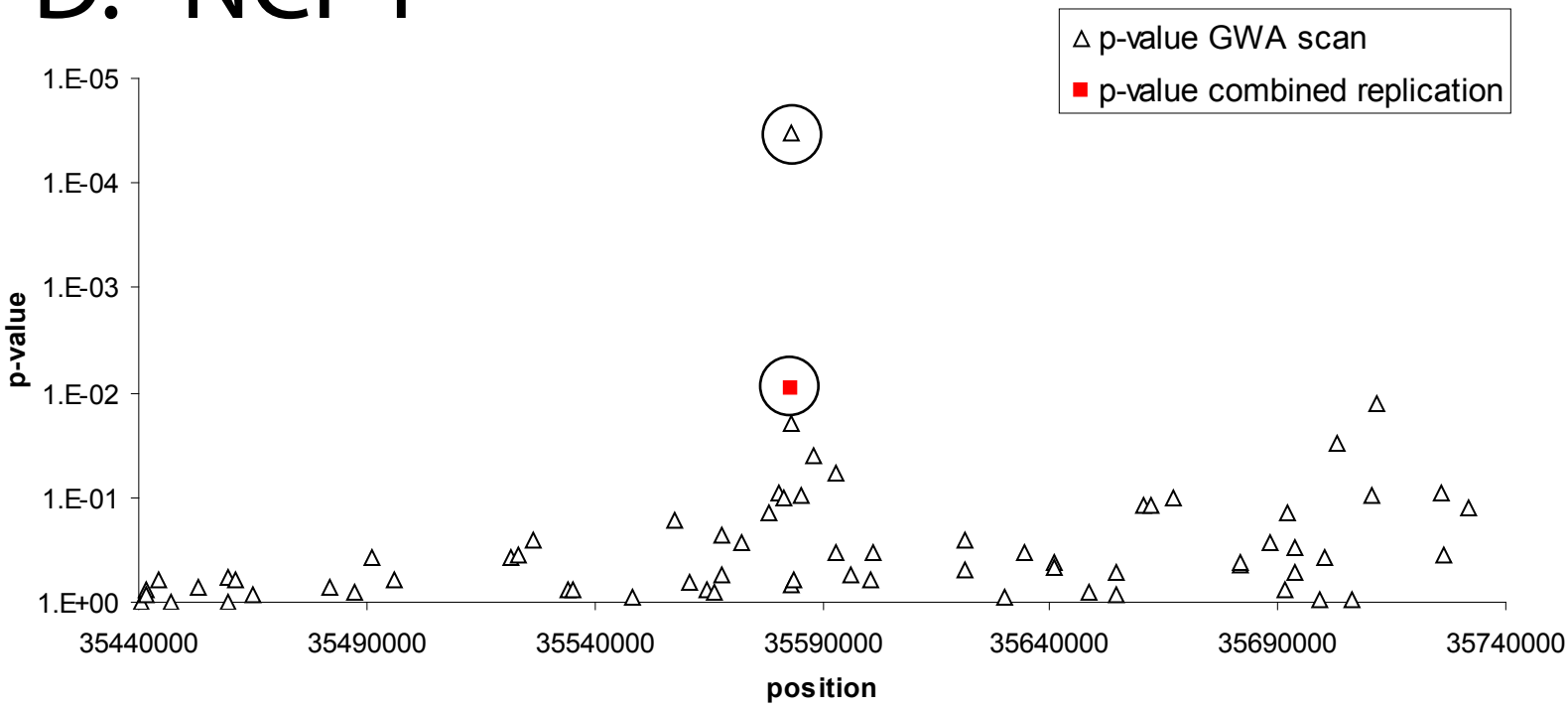

Chr10
63900k
64000k
64100k
64200k
64300k
Genotyped SNPs
Entrez genes
NM_199450
ZNF365: zinc finger protein 365 isoform B
NM_199451
ZNF365: zinc finger protein 365 isoform C
NM_199452
ZNF365: zinc finger protein 365 isoform D
NM_032804
C10orf22: hypothetical protein LOC84890
NM_000399
EGR2: early growth response 2 protein

C. PHOX2B


p-value combined replication
p-value GWA scan
1.E-07
1.E-06
1.E-05
1.E-04
1.E-03
1.E-02
1.E-01
1.E+00
p-value
41500000
41520000
41540000
41560000
41580000
41600000
41620000
41640000
41660000
41680000
41700000
position

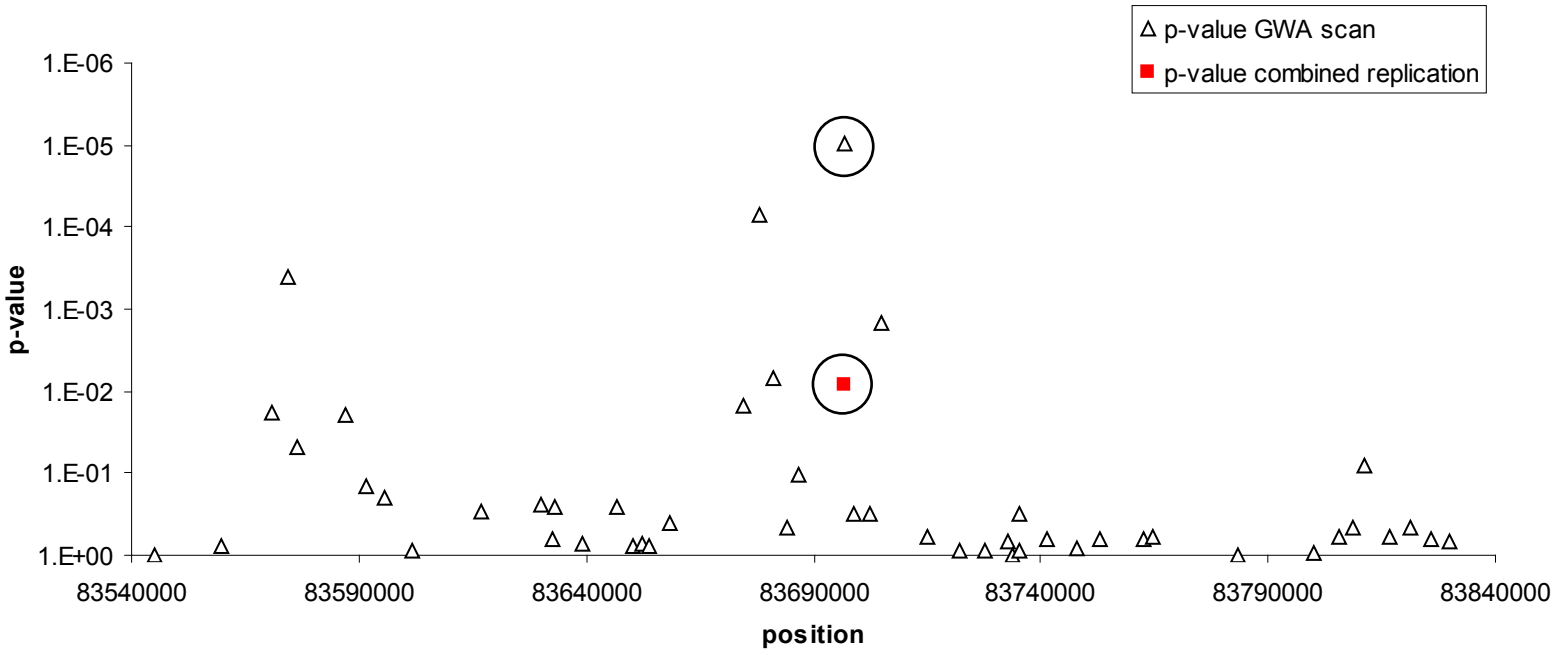

Chr4
Genotyped SNPs
Entrez genes
NM_014988
KIAA1102: hypothetical protein LOC22990
NM_003924
PHOX2B: paired-like homeobox 2b

D. NCF4


p-value GWA scan
p-value combined replication
1.E-05
1.E-04
1.E-03
1.E-02
1.E-01
1.E+00
p-value
35440000
35490000
35540000
35590000
35640000
35690000
35740000
position


Chr22
Genotyped SNPs
Entrez genes
NM_006860
RABL4: RAB
NM_002854
PVALB: parvalbumin
NM_207475
FLJ90680: hypothetical protein LOC400926
NM_000395
CSF2RB: colony stimulating factor 2 receptor
NM_178552
NM_000631
NCF4: neutrophil cytosolic factor 4 (40kD) isoform 1
NM_013416
NCF4: neutrophil cytosolic factor 4 (40kD) isoform 2

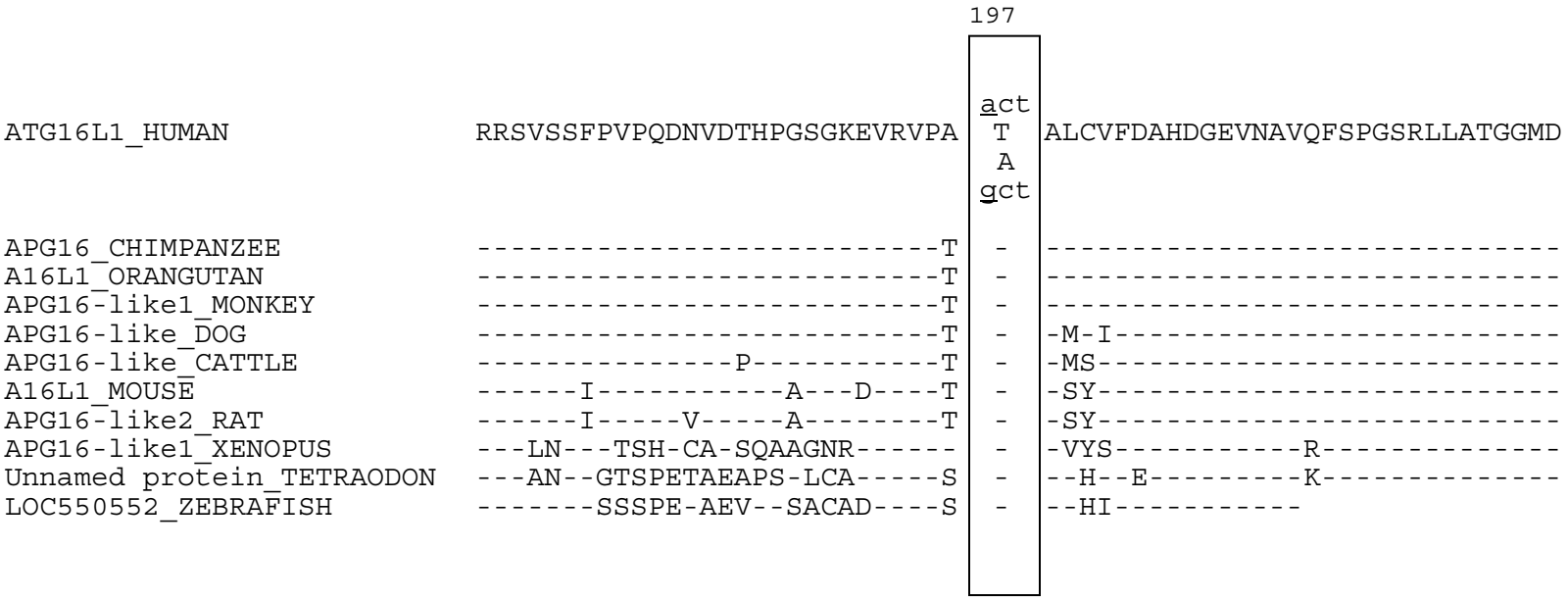
E. FAM92B


p-value GWA scan
p-value combined replication
1.E-06
1.E-05
1.E-04
1.E-03
1.E-02
1.E-01
1.E+00
p-value
83540000
83590000
83640000
83690000
83740000
83790000
83840000
position


Chr16
Genotyped SNPs
Entrez genes
NM_017740
ZDHHC7: zinc finger
NM_014732
KIAA0513: hypothetical protein LOC9764
NM_198491
FAM92B: hypothetical protein LOC339145

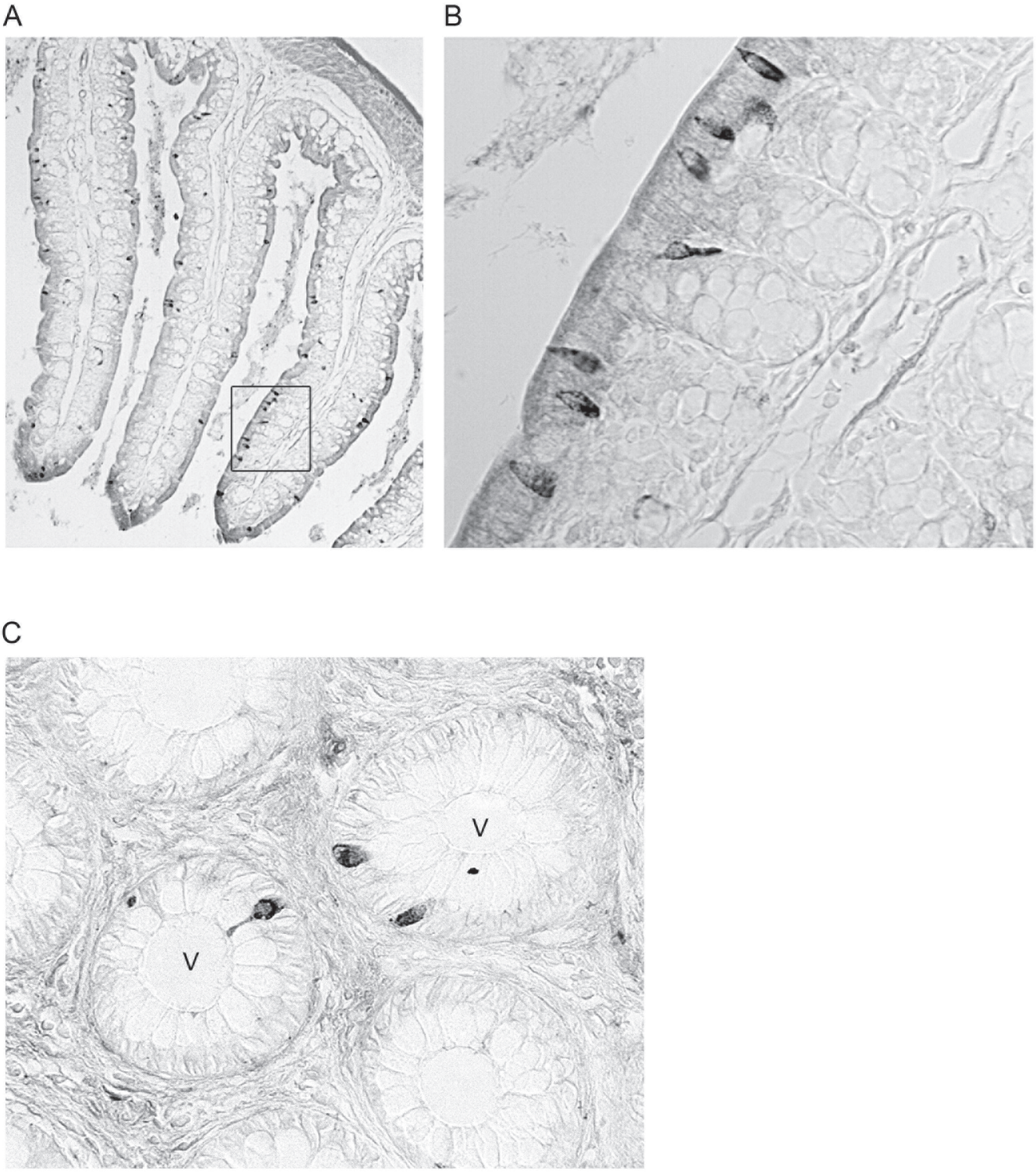
Supplementary Figure 1. Association and Linkage Disequilibrium patterns surrounding the ATG16L1, PHOX2B, 10q, NCF4, and FAM92B novel loci

Examination of the association and LD surrounding each of the five novel loci. The upper half of each panel shows a plot of association values (p values in negative logarithmic scale) for the SNPs from the GWA scan (open triangles) and from the combined replication studies (Replication Cohorts #1 and #2; red closed squares). In addition, in panel A, data from additional higher density mapping is presented (closed circles). The lower half of each panel shows the LD pattern in each region using the data from the GWA screen. The highest ranking association signal (circled on plot) is indicated by a red tick mark on the LD plot. The association signal decays on either side of the identified signal in these regions. Association signals are shown for regions surrounding the 5 novel CD susceptibility loci: (A) ATG16L1, 61 SNPs within a 500Kb region; (B) 10q21.1, 71 SNPs within a 500Kb region; (C) PHOX2B, 33 SNPs within a 100Kb region; (D) NCF4, 69 SNPs within a 300Kb region; and (E) FAM92B, 49 SNPs within 300Kb.



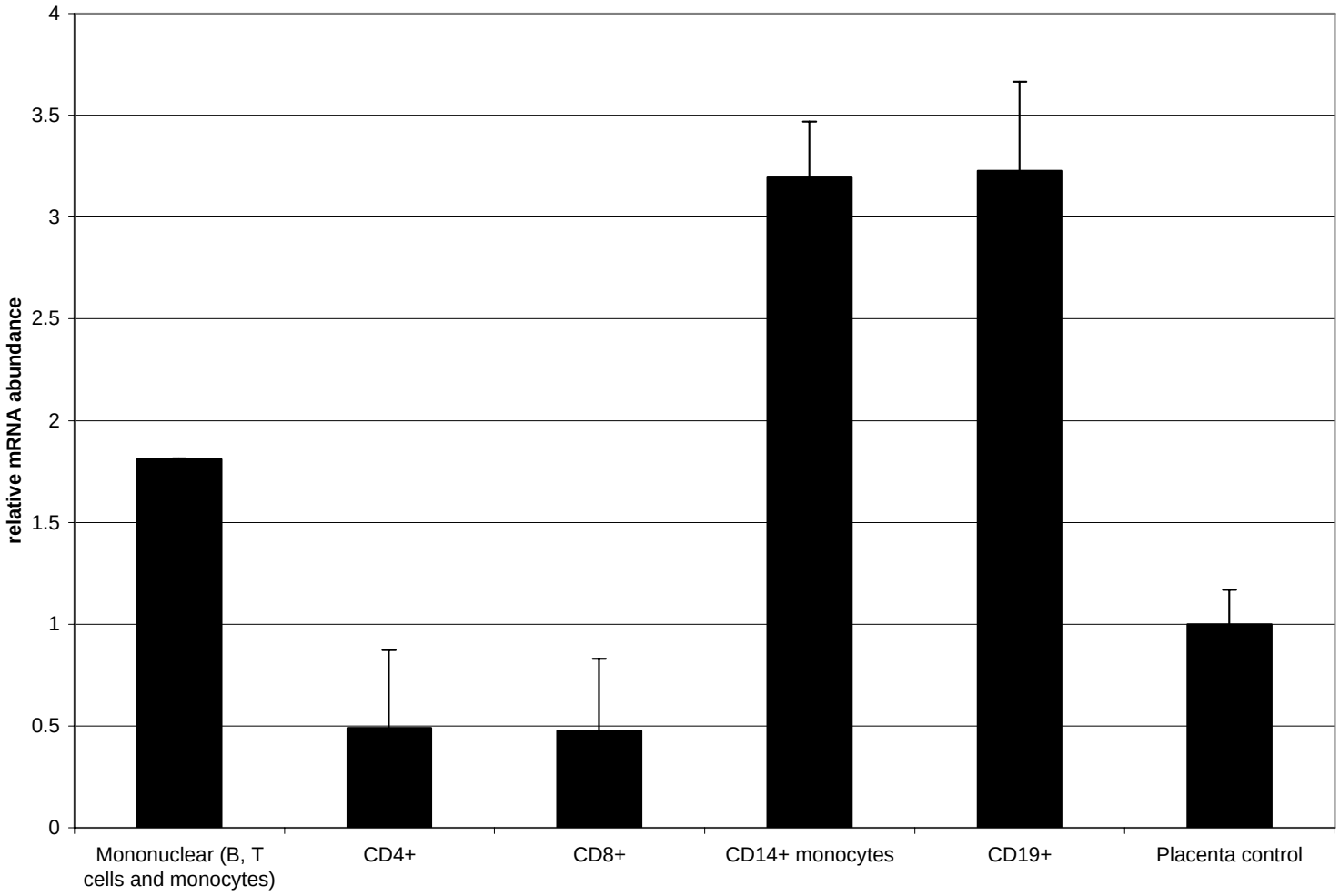
Supplementary Figure 2. Conservation of the threonine allele at position 197 of ATG16L1 gene

Sequence conservation between species. Amino acid sequence is shown as a single letter code (uppercase), where dashed lines indicate sequence identity with the human sequence shown. The box surrounds the associated coding variant at position 197 (Thr197Ala) and demonstrates conservation of the threonine across multiple species. The corresponding nucleotide triplet codon encoding this amino acid change is shown in lowercase letters where the variant base (corresponding to rs2241880) is underlined. The frequency of the Thr allele in the International HapMap database (www.hapmap.org) is 0.458, 0.725, 0.830, and 0.611 in the CEU, YRI, JPT, and HCB samples, respectively.



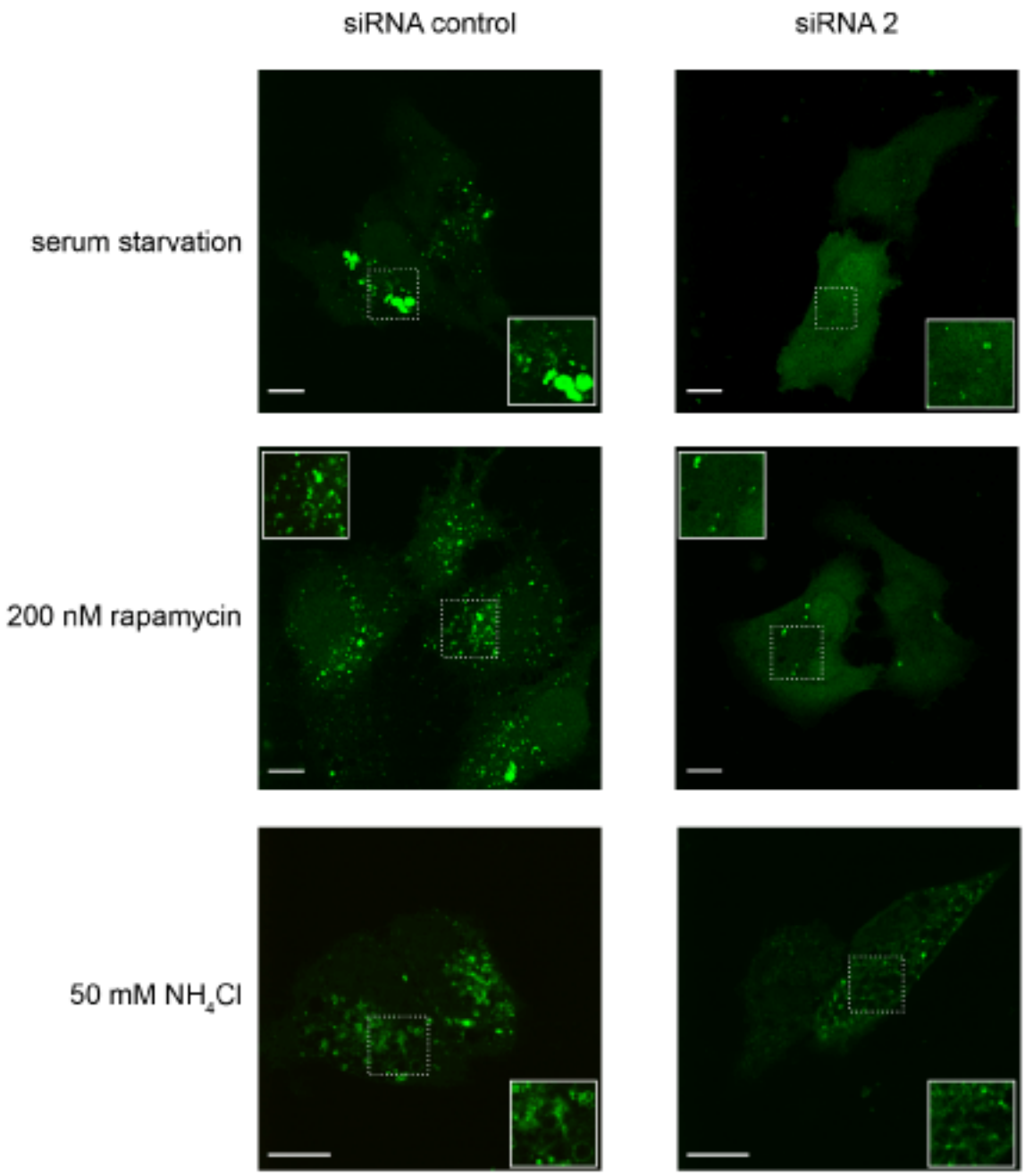
Supplementary Figure 3. PHOX2B expression is confined to a subset of gut cells in both mouse and human tissues

Phox2b expression was assayed using immunohistochemistry in both mouse ileum (A and B) and human colon (C). Paraffin-embedded sections were stained with specific antibodies and streptavidin HRP and developed with DAB. A - Positive cells are clearly visible as darkly stained cells within the epithelium of the villi, but are absent from the base of crypts. B - High magnification view of the area indicated by the black box in panel A. Phox2b positive cells are confined to the periphery of the villus and appear to be within the villar epithelium. C - A human colonic mucosal biopsy reveals Phox2b-positive cells within the epithelium, staining is confined to a specific subset of epithelial cells.



Supplementary Figure 4. Expression pattern of NCF4 in primary immune cells

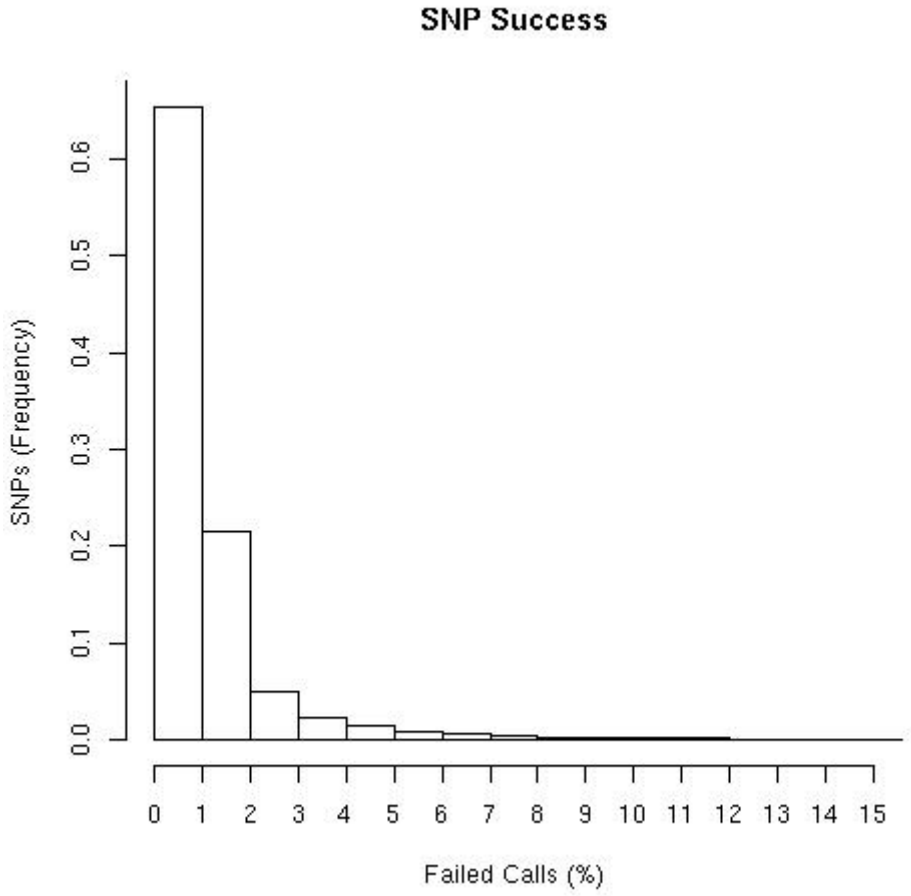
Quantitative real-time PCR was used to determine the expression patterns of NCF4 in human primary immune cells. In a resting human immune cell RNA panel (Clontech, CA) NCF4 showed a clear peak of expression in the CD14+ monocytes and in the CD19+ cells. Real-time quantitative RT-PCR reactions were performed in duplicate and the means plotted, error bars represent 1 standard deviation. Reactions containing no template RNA were performed to control for reaction contamination (water controls). Expression levels were normalized by comparison to GAPDH controls and arbitrary relative expression units plotted where placental RNA is equal to 1 (for cell lines, or primary cells respectively). All RNAs were isolated from resting cells in the absence of stimulation or activation.



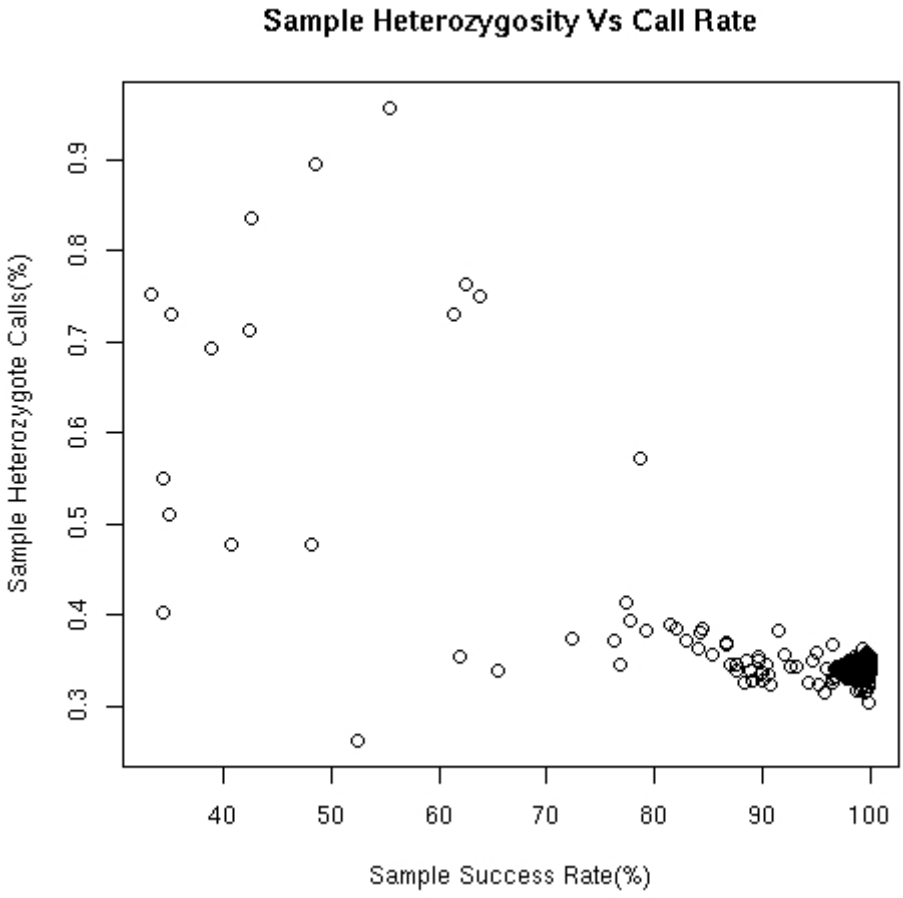
Supplementary Figure 5. Knockdown of ATG16L1 prevents induction of autophagy by classical stimuli

HeLa cells co-transfected with LC3-GFP plasmid and either control or ATG16L1-directed siRNA duplexes were subjected to classical autophagic stimuli. Control cells responded normally, with accumulation of numerous LC3-GFP vesicles after 24 hours of serum starvation (top panels) or rapamycin (middle panels) treatment. Cells subjected to ATG16L1 knockdown failed to accumulate LC3-GFP vesicles and retained a near-homogenous cytoplasmic pattern of LC3 distribution. When cells were treated to prevent lysosome fusion and proteolysis using ammonium chloride for 2 hours (bottom panels), control cells rapidly accumulated LC3+ vesicles, with membrane-bound LC3-GFP (visible as rings in confocal projection). In contrast, cells in which ATG16L1 had been knocked down exhibited a comparable accumulation of vesicles (visible as dark voids within the cytoplasmic LC3-GFP), but the majority of these did not appear to be bound by LC3+ membrane. Scale bars represent 10 μm.

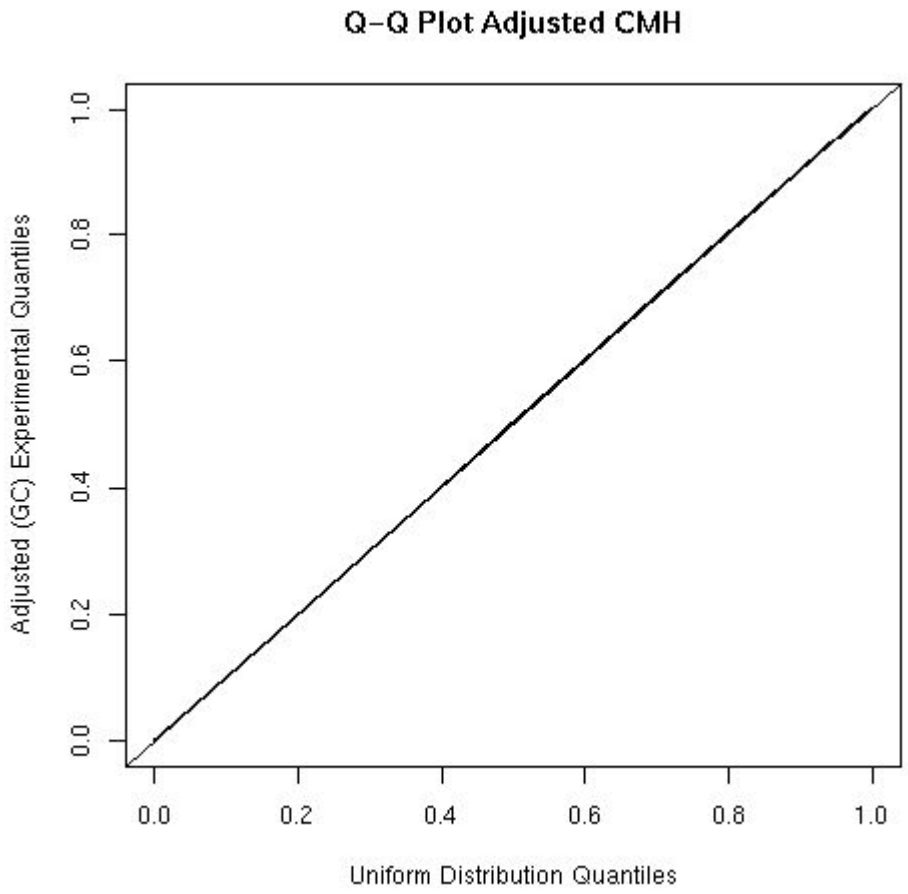
6A


SNP Success
SNPs (Frequency)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Failed Calls (%)

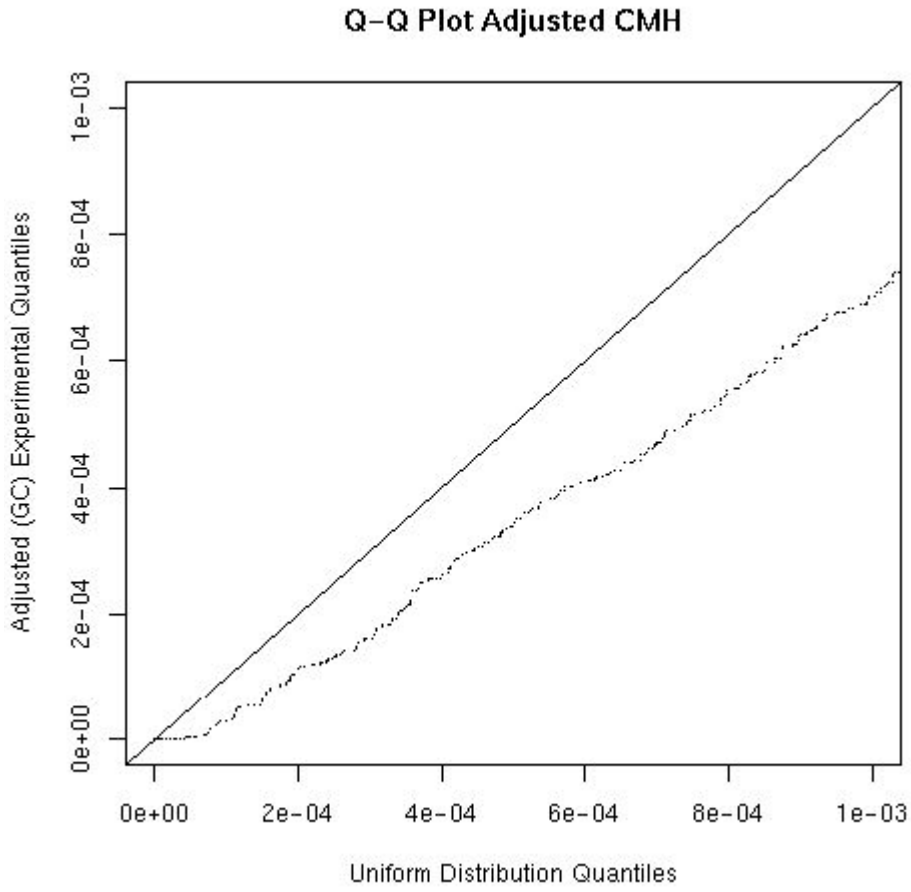
6B


Sample Heterozygosity Vs Call Rate
Sample Heterozygote Calls(%)
0.3
0.4
0.5
0.6
0.7
0.8
0.9
40
50
60
70
80
90
100
Sample Success Rate(%)

6C


Q-Q Plot Adjusted CMH
Adjusted (GC) Experimental Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
Uniform Distribution Quantiles
0.0
0.2
0.4
0.6
0.8
1.0

6D



Supplementary Figure 6: Quality control assessment of the GWA data
Panel 6A is a histogram of the genotyping call success rate; 6B is a plot of the heterozygosity as a function of genotyping call rate; 6C and 6D are Q-Q plots of the genomic control (GC) –adjusted Cochran-Mantel-Haenszel (CMH) p-values showing adherence to the null hypothesis for most of the distribution (6C) with strong deviation for the most extreme p-values (6D).