TITLE:
Environmental and genetic disease modifiers of haploinsuffciency of A20

Running title:
Hominin TNFAIP3 variants influence Mendelian disease

Authors:
Nathan W. Zammit[1,3], Paul E. Gray[4,5,], Owen M. Siggs[1,6], Jin Yan Yap[1], Amanda Russell[1], Daniele Cultrone[1,3], Joanna Warren[1], Stacey N. Walters[1], Robert Brink[1,3], David Zahra[1,3], Deborah L. Burnett[1,3], Velimir Gayevskiy[2,3], Andre E. Minoche[2,3], John B. Ziegler[4,5], Maria E. Craig[5,7], Melanie Wong[8], Paul Benitez-Aguirre[9], Juliana Teo[10], Mark J. Cowley[4,11], Marcel E. Dinger[1,12], Stuart G. Tangye[1,3], Catherine Burke[13], Tri G. Phan[1,3], Christopher C. Goodnow[1,3], Shane T. Grey[1,3]

Affiliations:
[1]Immunology Department, [2]Kinghorn Center for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW, 2010 Australia
[3]St Vincent's Clinical School, Faculty of Medicine, University of New South Wales Sydney, NSW 2145, Australia
[4]Department of Immunology and Infectious Diseases, Sydney Children's Hospital, Randwick NSW 2031, Australia
[5]School of Women's and Children's Health, University of New South Wales, Sydney, NSW 2010, Australia
[6]Department of Ophthalmology, Flinders University, Bedford Park, SA 5042 Australia
[7]Institute of Endocrinology and Diabetes, [8]Department of Allergy and Immunology, [9]Endocrinology Department, [10]Haematology Department, The Children's Hospital at Westmead, NSW 2045, Australia
[11]Children's Cancer Institute, Lowy Cancer Centre, UNSW Sydney, Kensington, NSW, Australia
[12]School of Biotechnology and Biomolecular Sciences, UNSW Sydney, NSW 2052, Australia
[13]School of Life Sciences, University of Technology Sydney, Sydney, NSW 2007, Australia

ABSTRACT

Monogenic diseases can often manifest diverse clinical phenotypes and cause diagnostic dilemmas. While monoallelic loss-of-function variants in *TNFAIP3* (Haploinsufficiency of A20; HA20) cause a highly penetrant autoinflammatory disease, the variable expressivity suggest a role for additional genetic and environmental disease modifiers. Here, we identify critically ill children who inherited a family-specific *TNFAIP3* deletion from one of their otherwise healthy parents. Each of the probands also inherited *in trans* a subtle loss-of-function I207L *TNFAIP3* variant that is common in Oceania, originally introgressed from Denisovans. Modelling this compound heterozgous state in mice under specific pathogen free conditions demonstrated a reduced threshold to break immune tolerance. Exaggerated immune responses were precipitated by inheriting the two genetic hits on the *TNFAIP3* checkpoint coupled with increasing the microbial challenge to immune tolerance, either by co-housing with “pet store” mice carrying a wild microbial burden or by transient dietary exposure to a chemical that diminishes the intestinal mucin barrier separating gut microbes from immune sensing systems. These data illuminate second-hit genetic and environmental modifiers contributing to complex inflammatory and autoimmune disease. Increased mechanistic understanding of the presence and contribution of disease modifiers will aid diagnostic and prognostic patient stratification and potentially reveal novel therapeutic opportunities.

53 **INTRODUCTION**

54 *TNFAIP3* (tumor necrosis factor, alpha-induced protein 3) encodes A20, a highly
55 conserved and evolutionarily ancient protein that serves as a key immune tolerance
56 checkpoint by negatively regulating tissue inflammation and immune responses by
57 inhibiting NF-κB signaling (1-4). A20 inhibits NF-κB signalling though catalytic and
58 non-catalytic mechanisms. A20's ovarian tumor domain (OTU) and Zinc finger 4 region
59 exert K63-deubiquitinating and K48-ubiqutinating activity, respectively, and co-operate
60 to inactivate key signaling substrate proteins and target them for proteasomal
61 degradation (5-8). In contrast, A20's Zinc finger 7 binds to linear ubiquitin chains
62 disrupting IKK activation (9, 10). *TNFAIP3* transcription is induced by NF-κB (11-13)
63 and A20's inhibitory activities are activated by IKKβ-mediated phosphorylation at Ser
64 381 (9, 14, 15), forming a negative feedback loop on NF-κB signaling.

65

66 Common variants at the *TNFAIP3* locus are associated with multiple polygenic
67 autoimmune and inflammatory diseases (16), including type 1 diabetes, rheumatoid
68 arthritis, systemic lupus erythematosus, psoriasis and Crohn's disease highlighting the
69 importance of *TNFAIP3* in human health and disease. In contrast, haploinsufficiency
70 of A20 (HA20) caused by *TNFAIP3* deletion or protein-truncating variants causes an
71 autosomal dominant, childhood-onset inflammatory disease, reminiscent of Behcet's
72 disease, with apparent complete penetrance (17). Subsequent findings have reported
73 HA20 in patients with diverse autoimmune and inflammatory dysregulatory conditions
74 including autoimmune lymphoproliferative syndrome (ALPS), systemic lupus
75 erythematosus, type 1 diabetes, hepatitis and interstitial lung disease (18-24).

76

77 An emerging feature of HA20 is its association with diverse clinical presentations,
78 phenotypes, and disease severity, even in individuals carrying the same variant, as
79 well as a roughly two times higher incidence in females then men (19, 23-25). The
80 recognition of diverse phenotypes, age of onset and gender bias, suggests a
81 significant role for environmental and or genetic disease modifiers in patients with
82 HA20. Such interactions have been described for Mendelian diseases, including the
83 identification of genes that can modify cystic fibrosis phenotypes in association with
84 mutations in the *CFTR* gene (26), and both gene and environmental factors that
85 influence Huntington's disease progression (27). However, few have been identified
86 in the context of monogenic immune dysregulatory conditions.
87
88 Here, we identify critically ill children who inherited a *TNFAIP3* deletion from one of
89 their otherwise healthy parents, challenging the hypothesis that A20 deficiency alone
90 is always fully penetrant for disease. We now explain these apparently contradictory
91 findings by providing evidence from mouse models for two additional disease modifiers
92 contributing to HA20 in these patients. The first disease modifier is an inherited archaic
93 *TNFAIP3* genetic variant that causes a subtle loss of A20 function, while the second
94 disease modifier came in the form of an environmental challenge that altered
95 microbiome composition. HA20 is a new disease and increased mechanistic
96 understanding of the presence and contribution of disease modifiers will aid diagnostic
97 and prognostic patient stratification and potentially reveal novel therapeutic
98 opportunities.

RESULTS

Here, we studied two families, Family 1 of mixed Scottish-Fijian/Samoan ancestry and Family 2 of European and New Zealand Māori ancestry (Figure 1A), each with a child with early-onset inflammatory and autoimmune manifestations (Figure 1B, Supplementary Figure 1A,B). Both Child 1 (Family 1) and Child 2 (Family 2) presented with inflammatory bowel disease (IBD), autoantibodies, and cytopenia. Child 1 also exhibited ocular inflammation, subclinical thyroid disease and rhinitis, whereas Child 2 also presented with mouth and genital ulcers, arthralgia/arthritis and type 1 diabetes. Whole genome sequencing of the parent-proband trios revealed each family harbored a unique deletion at the *TNFAIP3* locus, present in both the proband and one unaffected parent respectively. The deletion in Family 1 spanned the entire *TNFAIP3* locus (g.137683250-139345585del; Figure 1C) and was confirmed by CGH array, while Family 2 harboured a 2bp frameshift deletion predicting a premature stop codon truncating the normal protein sequence (p.(N371Sfs*17)) (Figure 1D, Supplementary Figure 2). However, neither of the families had a history of childhood-onset inflammatory disease, despite the fact that an unaffected parent in each pedigree aged 59 and 40 respectively, carried the same heterozygous *TNFAIP3* loss of function allele as the affected proband (Figure 1A). Analysis of human sequence data from people without early onset disease (gnomAD v2.1.1) revealed only 2 from 141,456 individuals with heterozygous protein-truncating *TNFAIP3* mutations, compared to an expected rate of 32 individuals based on coding region size and composition (pLI=1, LOEUF=0.2) (28). Studies in vertebrate model organisms also reveal that HA20 does not always drive a disease phenotype, as mice and zebrafish heterozygous for *Tnfaip3* deletion present as healthy when housed under specific pathogen free (SPF) conditions (1, 3). This suggested that additional, but hitherto unknown, disease

124 modifiers cooperate with *TNFAIP3* haploinsufficiency in these families to drive their
125 disease phenotypes.
126
127 To explore the potential of additional genetic modifiers, we searched for *TNFAIP3*
128 variants *in trans* with each deletion. In both probands we identified the same pair of
129 missense variants *in cis*, p.(T108A) and p.(I207L) (Figure 1D), which formed part of a
130 larger haplotype consisting of multiple coding and non-coding variants (15, 29). Both
131 missense variants are within the OTU domain of A20 (Figure 1E-G), and were rare
132 within public variant databases (gnomAD v2.1.1 allele frequency 0.0001169-
133 0.0001202). We have previously shown, these variants are enriched in populations in
134 Oceania (15), representing 15-45% of all alleles in Polynesia (e.g. Vanuatu), ~22% of
135 alleles in Martu Australians, 31-48% of *TNFAIP3* alleles in people of the eastern
136 Indonesian islands (e.g. Flores), and 32-100% of alleles in Papuans. Analysis of the
137 genetic history of this haplotype revealed it was inherited by anatomically modern
138 humans from Denisovans ~50 kya (15, 29). The *TNFAIP3* Denisovan haplotype (hap)
139 contains other non-coding variants found not to impact *TNFAIP3* expression (15). In
140 addition, I207 is predicted to be the functional coding allele, as in contrast to T108,
141 I207L is invariant across most jawed vertebrates (Figure 1F) and has a higher *in-silico*
142 deleterious score (CADD score 23.2; (15)). In transfected cells expressing *TNFAIP3*
143 cDNA, the T108A;I207L substitutions decreased A20 phosphorylation by IKKβ to 80%
144 of wild-type, creating a subtle loss of NF-κB inhibitory function (15). In both families,
145 the T108A;I207L Denisovan haplotype was inherited from one healthy parent and the
146 *TNFAIP3* deletion from the other healthy parent (Figure 1A). Hence, it was conceivable
147 that co-inheritance of *TNFAIP3* haploinsufficiency and a subtle loss of function

148 Denisovan allele explained disease only in the child with both genetic hits in compound
149 heterozygous configuration.
150
151 To test if coinheritance of the two alleles had additive biochemical affects we
152 performed Western blot analysis of peripheral blood mononuclear cells (PBMCs) from
153 healthy donors, the heterozygous parents and the affected patients. A20 protein levels
154 were decreased in each of the deletion-carrying parents compared to healthy donors
155 (Hd) (Figure 2A, B & Supplementary Figure 3. A, C). Neither *TNFAIP3* deletion 1
156 (+/del1), spanning the entire *TNFAIP3* locus, or deletion 2, a frameshift deletion
157 p.(N371Sfs*17), encoded proteins that could be detected by the antibody probe raised
158 against the C-terminus region (aa 440-790) of A20. Indeed, mRNA analysis of
159 *TNFAIP3* levels using probes spanning exons 2 and 4 show a dose dependent
160 reduction for individuals in family 1 harboring a heterozygous deletion of the entire
161 *TNFAIP3* locus compared to 15 healthy donors (Figure 1C). Heterozygous carriers of
162 the I207L allele (+/hap) exhibited similar A20 protein levels to healthy donors, with
163 probands (hap/del) exhibiting levels similar to their +/del parent (Figure 2A, B &
164 Supplementary Figure 3A, C). Active IKKβ also phosphorylates another NF-κB
165 inhibitory protein, IκBα, which then undergoes K48-linked ubiquitination and
166 degradation to release active NF-κB (30, 31). Western blot analysis revealed levels of
167 IκBα were greatly decreased in PBMCs from the del/hap probands in both families
168 even under non-stimulated conditions (Figure 2A, B; Supplementary Figure 3B, D).
169 For family 1, PBMCs from the healthy parent with the *TNfAIP3* deletion (+/del)
170 exhibited more rapid loss of IκBα following TNF stimulation compared to PBMCs from
171 healthy donors (Figure 2A; Sup. Figure 3B), as previously reported for HA20 (17).
172 However, this effect was less than the del/hap proband from this family (Figure 2A;

173 Sup. Figure 3B). In family 2, the healthy +/del parent exhibited greatly decreased IκBα
174 similar to the del/hap proband (Figure 2B; Sup. Figure 3D).
175
176 RT-PCR analysis of PBMCs was used to measure mRNA encoding NF-κB induced
177 inflammatory genes (30) before and following TNF stimulation. These target genes
178 were more strongly induced in cells from del/hap probands than in PBMCs from
179 *TNFAIP3*$^{+/+}$ donors. Cells from +/del individuals exhibited an intermediate phenotype,
180 while +/hap cells showed a subtle hyper-inflammatory response (Figure 2C, D).
181 Collectively, these data are consistent with an additive biochemical effect of
182 coinheriting a *TNFAIP3* null/truncating and *TNFAIP3* hypomorphic allele, but are not
183 conclusive because of the small number of donors with the relevant genotypes.
184
185 Given the limited availability of matched +/del and hap/del individuals to make
186 conclusions about additive effects, we generated two mouse lines by Cas9/CRISPR
187 gene editing on the genetically homogeneous C57BL/6 background, transmitting
188 either a *Tnfaip3* truncating mutation (p.Ala240Valfs*7, referred to as del) or a *Tnfaip3*
189 T108A;I207L allele (hap) (15) (Supplementary Figure 4A, B). Next we crossed both
190 lines so that offspring would be wild-type, heterozygous truncating (+/del),
191 heterozygous T108A;I207L (+/hap), or compound heterozygous (hap/del) to model the
192 human family members (Figure 3A). *Tnfaip3* +/hap, +/del and hap/del mice reared in
193 SPF conditions exhibited normal weights at weaning and at 8 weeks of age (Figure
194 3B; Supplementary Figure 4C, D). Heterozygous loss of full length A20 due to the
195 truncating allele was confirmed by a decrease to 50% of wild-type A20 protein levels
196 in isolated +/del and hap/del thymocytes (Figure 3C, D). Western blot analysis of TNF-
197 stimulated thymocytes revealed decreased IκBα protein levels even under non-

198 stimulated conditions for compound heterozygous hap/del cells that was not observed
199 for thymocytes from +/hap or +/del thymocytes (Figure 3C, E). *Tnfaip3* +/del cells
200 exhibited a trend to increased stimulus-dependent loss of IκBα compared to +/hap and
201 WT cells (Figure 3C, E). These findings highlight an additive role for +/del and +/hap
202 *Tnfaip3* alleles in restraining NF-κB inhibitory activity.
203
204 In contrast to the exaggerated biochemical response in the compound heterozygous
205 mice, flow cytometric analysis revealed comparable increases in the proportions of
206 $CD44^{hi}CD62^{low}$ effector memory CD8 and CD4 T cell subsets in the spleen (Figure 3F)
207 and mesenteric lymph nodes (MLN, Figure 3G) of *Tnfaip3* +/del and hap/del mice
208 compared to +/hap and wild-type littermates (Supplementary Figure 5A-F). *Tnfaip3*
209 +/del and hap/del mice also exhibited a comparably increased proportion of FOXP3+
210 Tregs among CD4 T cells (Figure 3H, I), and increased number of B cells in the spleen
211 (Supplementary Figure 5G). However, only hap/del mice exhibited significantly
212 increased numbers of germinal center B cells in the spleen and MLN under SPF
213 conditions (Supplementary Figure 5I, J).
214
215 A major clinical phenotype in HA20 patients included gut pathology (Supplementary
216 Figure 1A). Interestingly, mucosal inflammation is a unifying symptom of human HA20
217 disease (24, 32) and spontaneous gut inflammation observed in *Tnfaip3* gene-targeted
218 mice is driven by exaggerated responses to gut microbiota (33). Analysis of colon
219 sections from SPF *Tnfaip3* hap/del mice, revealed immune infiltrates with a mild
220 clinical score similar to *Tnfaip3* +/del mice, which was higher than that observed
221 in *Tnfaip3* +/hap or wild-type littermates (Figure 3J, K).
222

223 The results above indicated that *Tnfaip3* haploinsufficiency (+/del) in mice causes
224 subclinical immune dysregulation that is not markedly increased by co-inheriting the
225 hypomorphic I207L allele when the microbial flora is highly controlled in an SPF facility.
226 To better replicate natural microbial stimulation of the immune system, we co-housed
227 the *Tnfaip3* mutant mice with "dirty" mice harboring a diverse microbiome. Cohousing
228 mice with non-laboratory wild mice has been shown to facilitate the horizontal transfer
229 of microbiota, maturing the immune system (34, 35). Deep sequencing of bacterial
230 16S rRNA of fecal pellets from the co-housed mice showed horizontal transfer
231 resulting in a more diverse gut microbiota compared to their SPF counterparts,
232 consisting of an increase in the relative abundance of Prevotella, Odoribacter,
233 Helicobacter, Muscispirillum and a decrease in Akkermansia. However, no discernible
234 difference in principle microbial components was observed among co-housed *Tnfaip3*
235 hap/del, +/del, +/hap and +/+ animals (Figure 4A; Supplementary Figure 6A, B).
236 Cohousing did not adversely affect the weight or survival of *Tnfaip3* wild-type, +/hap,
237 +/del or hap/del mice (Supplementary Figure 6C), but increased the percentage of
238 $CD44^{hi}$ effector/memory CD4 and CD8 T cells in the blood within 1 week of cohousing
239 (Supplementary Figure 6D, E). Following 8 weeks of cohousing, clinical scores for
240 colon inflammation tended to be higher in *Tnfaip3* hap/del mice than +/del animals,
241 with marked mononuclear infiltrates and goblet cell hyperplasia (Figure 4B, C). In the
242 MLN draining the gut, the proportion of effector/memory cells among CD8 and CD4 T
243 cells was increased in +/del mice and further increased in hap/del compound
244 heterozygotes (Figure 4D, E). A similar finding was observed when measuring total
245 CD44+ CD8 T cells (Supplementary Figure 5D), and CD4 central memory T cells in
246 MLNs (Supplementary Figure 5E, F). Cohousing with dirty mice did not segregate Treg
247 frequencies between *Tnfaip3* +/del and hap/del genotypes (Figure 4F, G) but

248 significantly decreased mean FOXP3 protein levels per cell in splenic Tregs of *Tnfaip3*
249 hap/del mice compared to +/del mice (Figure 4H). In cohoused mice, percentage of
250 splenic germinal center B cells was progressively increased according to *Tnfaip3*
251 genotype consistent with additive effects of the del and hap mutations (Supplementary
252 Figure 5 H-J).
253
254 To test a different environmental stressor, we next challenged *Tnfaip3* hap/del mice
255 maintained in SPF conditions by adding 2.5% dextran sodium sulfate (DSS) to drinking
256 water. DSS thins the intestinal mucin barrier that normally promotes bacterial tolerance
257 by physically separating most gut microbiota from the intestinal epithelium. DSS was
258 given for 5 days, then removed (day 0) and weight loss and recovery monitored for 2
259 weeks. Weight loss was exaggerated and recovery delayed in *Tnfaip3* +/del animals
260 and these effects were worsened in hap/del compound heterozygous animals (Figure
261 4I). Only *Tnfaip3* hap/del mice exhibited significant weight loss over the test period
262 compared to wild-type littermates. Two weeks post-DSS removal, colons were
263 collected and assessed for inflammatory cell infiltrate, erosion and ulceration,
264 hyperplasia and submucosa oedema and goblet cell and crypt loss. *Tnfaip3* hap/del
265 mice exhibited the highest clinical scores (Figure 4J, K). DSS challenge caused an
266 increase in effector/memory $CD44^{hi}$ CD8 T cells and $FOXP3^{+}$ Tregs for all genotypes
267 assessed (Supplementary Figure 7A-D). In Tregs, *Tnfaip3* hap/del mice treated with
268 DSS exhibited a significant decrease in mean FOXP3 protein per cell compared to
269 wild-type littermates treated with DSS, while +/del animals tended to be intermediate
270 (Figure 4L, M). DSS treatment also increased the frequency of neutrophils in the
271 spleen and MLNs, with *Tnfaip3* hap/del mice exhibiting higher frequency of splenic

272 neutrophils compared to +/del alone and wild-type littermates (Supplementary Figure
273 7E, F).

274 **DISCUSSION**

275 The findings here reveal how rare and common inherited variants in an important
276 immune tolerance checkpoint interact with environmental stressors to precipitate
277 autoimmune and inflammatory disease. Rare deletions in *TNFAIP3* have been
278 considered to cause a fully penetrant autosomal dominant, childhood-onset
279 inflammatory/autoimmune disease by haploinsufficiency (17). However, HA20
280 patients present with diverse clinical characteristics and variable age of onset
281 suggesting a role for additional disease modifiers. Here we describe two children with
282 clinical HA20 syndrome who inherited a family-specific, heterozygous inactivating
283 *TNFAIP3* deletion from one well parent and a subtle loss-of-function *TNFAIP3* allele
284 that is common in Oceania, from their other healthy parent (15, 29). Ex-vivo analysis
285 of patient cells demonstrates increased inflammation and NF-κB signaling in
286 asymptomatic deletion carriers, which is potentiated in symptomatic probands. This
287 potentiation was replicated in mice rendered compound heterozygous for a deletion
288 and the common Oceanian allele, precipitated by exposing the mice to diverse gut
289 microbes or damaging the mucin barrier between gut microbes and the body's microbe
290 response systems. The families described here demonstrate that contrary to current
291 consensus (18-22), *TNFAIP3* haploinsuffucuency does not always cause a fully-
292 penetrant childhood-onset inflammatory disease, and that *TNFAIP3*-associated
293 inflammatory disease exemplifies interaction between environmental triggers and
294 common and rare inherited alleles.

295

296 Here we show for the first time that some people carrying a heterozygous TNFAIP3
297 deletion/truncating allele do not develop inflammatory/autoimmune disease. Why does
298 this occur? The animal experiments here support two explanations that are not

299 mutually exclusive. First, asymptomatic carriers may lack second genetic hits – single
300 or multiple common polymorphisms - that further compromise the *TNFAIP3* immune
301 tolerance checkpoint. The 'gene dose' effect observed for *TNFAIP3* variants here may
302 be analogous to the potentiation of rare deletions in RBM8A by common *in trans*
303 hypomorphic SNPs in the causation of the thrombocytopenia with absent radii (TAR)
304 syndrome (36). By contributing to compound heterozygous inheritance of HA20, the
305 Denisovan-derived I207L (*Tnfaip3* hap) is a rare example of a variant with relevance
306 both at a population and Mendelian level, analogous to protection from severe
307 Falciparum Malaria afforded by heterozygosity for sickle cell variants (37).
308
309 Second, these genetically predisposed, asymptomatic individuals may not have had
310 as much pressure applied to the checkpoint by environmental microbes. These
311 concepts build upon Archibald Garrod's prototypic recessive "inborn error of
312 metabolism", alkaptonuria (AKU) (38), and the more common recessive disease
313 phenylketonuria (PKU) (39), where the genetic dose of mutations inactivating enzymes
314 for phenylalanine and tyrosine metabolism cooperates with the dietary load of these
315 amino acids. Above a certain threshold, over time, accumulating metabolic
316 intermediates form toxic polymers on collagen (alkoptonuria) or unmetabolized
317 phenylalanine blocks transporters at the blood-brain barrier leading to developmental
318 delay. In the case of PKU, understanding the gene/environment cooperation enables
319 disease prevention by limiting dietary phenylalanine. For people with an underlying
320 gene deficit in the *TNFAIP3* immune tolerance checkpoint, it will be important to
321 recognise these individuals early and identify ways to minimise microbial stimuli that
322 take them over a threshold into inflammatory or autoimmune disease.
323

324 Exposing *Tnfaip3* hap/del mice to ‘dirty mice’ with ‘wild’ microbial flora or transiently
325 introducing DSS in the drinking water (40), revealed a significantly worse outcome for
326 *Tnfaip3* hap/del mice compared to single variant mice. A thick inner mucin layer
327 secreted by intestinal goblet cells normally serves a key role in tolerating microbial
328 flora in the gut lumen. This is achieved by physically separating them from bacterial-
329 sensing receptors on and within intestinal epithelial cells and lamina propria
330 macrophages, notably Toll-like receptors (TLR) and NOD-like receptors (NLR) that
331 activate NF-κB. Oral DSS rapidly disrupts the mucin barrier to increase bacterial
332 contact with the epithelium (41, 42). The resulting stimulation of TLR- and NLR-
333 induced NF-κB activation will increase pressure on the *TNFAIP3* checkpoint to
334 dampen NF-κB signaling and maintain tolerance to the bacteria, providing an
335 explanation for the additive effects of compound heterozygous hap/del *TNFAIP3*
336 mutations on DSS pathology. Consistent with this interpretation, complete deletion of
337 both TNFAIP3 alleles selectively in murine intestinal epithelial cells cooperates with
338 DSS treatment and gut flora to cause a similar inflammatory bowel disease (43).
339
340 As opposed to dietary compromise to the mucin barrier, a number of the bacterial
341 species that were expanded in the intestine of the co-housed mice, have specialized
342 mechanisms to penetrate the mucin barrier and adhere directly to intestinal epithelium.
343 Gram-negative Prevotella species of the Bacteroidetes phylum and segmented
344 filamentous bacteria predispose to autoimmune disease (44, 45) by mechanisms that
345 depend upon adherence to intestinal epithelium (46, 47). Mucispirillum colonizes the
346 mucin layer to provoke inflammatory bowel disease in mice lacking NOD2 and CYBB
347 NADPH oxidase (48). Helicobacter hepaticus provokes colitis in IL-10 deficient mice
348 (49). These expansions in bacterial species coincided with a loss of Akkermansia that

349 exhibit anti-inflammatory properties that maintain the mucin barrier (50, 51). In mice
350 with specific pathogen free flora, homozygous loss of TNFAIP3 in intestinal epithelial
351 cells and in intestinal macrophages results in a vicious cycle of microbe-induced
352 overproduction of TNF by macrophages and intestinal epithelial cell hyperproliferation
353 and apoptosis, depleting goblet cells and the mucin barrier (52). Collectively, these
354 results illustrate how genetics, diet and gut microbial flora cooperate to exceed
355 tolerance thresholds.
356
357 A significant loss of FOXP3 MFI was observed in *Tnfaip3* hap/del Tregs following
358 cohousing and DSS treatment that was not observed for single variant mice,
359 potentially explaining their segregating phenotypes. Treg activation and expansion is
360 NF-κB dependent via the TCR-CARD11 pathway (53-55). This is further evidenced by
361 expanded Treg populations observed in the *Tnfaip3* variant mice (15). However,
362 chronic inflammation and persistent NF-κB activation can drive FOXP3 instability (56,
363 57) by preventing FOXP3 hypomethylation, which requires CD28/NF-κB deprivation
364 in developing Tregs (57). These data suggest that the biallelic addition of +/hap and
365 +/del, potentiates NF-κB activation under inflammatory conditions, down-regulating
366 FOXP3 expression. Indeed western blot analysis of TNF stimulated thymocytes show
367 an additive effect for *Tnfaip3* +/hap and +/del variants on the NF-κB inhibitor, IκBα. As
368 A20 haploinsufficiency lowers immune stimulatory thresholds, a loss of FOXP3
369 expression would likely be detrimental contributing to clinical inflammatory conditions
370 (58-61).
371
372 The experimental approaches deployed here possess real potential in the study of
373 both HA20 and polygenic inflammatory bowel disease (IBD). The failure of mouse

374 avatars representing the genetics of human monogenic inflammatory disease to
375 express a clinical phenotype unless specifically challenged is not unique, with *Lrba*
376 gene-deficient animals requiring DSS stimulation to develop clinical disease (62). The
377 models of colitis presented here associated with mice having a more 'wild-type'
378 intestinal flora contributes to growing knowledge for human IBD where the search for
379 pathogenic mechanisms is increasingly focused on changes to the intestinal
380 microbiota (63).
381
382 In summary, we demonstrate for the first time, incomplete penetrance associated with
383 deletions in *TNFAIP3*. The Oceanic allele that contains the functional p.I207L allele
384 represents a unique example of a common variant contributing to a Mendelian
385 inflammatory disease, and has implications for genomic filtering strategies, which
386 might remove such variants. Furthermore, *TNFAIP3* T108A;I207L which was
387 adaptively introgressed into Oceanian populations and can potentially provide
388 resistance to infection (15), may represent the first example of an immune variant with
389 a population protective benefit, that is Mendelian deleterious. It remains to be seen
390 whether this same variant influences immunity and inflammation in other non-
391 Mendelian diseases in the large number of individuals who carry it today. Finally, our
392 data highlight the significance of ancestry in rare disease penetrance, and the value
393 of studying inherited disease in non-European populations, which are poorly
394 represented in current clinical and population genetic studies.

395 **REFERENCES**
396
397 1. Lee EG, Boone DL, Chai S, Libby SL, et al. Failure to regulate TNF-induced NF-kappaB and cell
398 death responses in A20-deficient mice. Science. 2000;289(5488):2350-4.
399 2. Boone DL, Turer EE, Lee EG, Ahmad RC, et al. The ubiquitin-modifying enzyme A20 is
400 required for termination of Toll-like receptor responses. Nat Immunol. 2004;5(10):1052-60.
401 3. Cultrone D, Zammit NW, Self E, Postert B, et al. A zebrafish functional genomics model to
402 investigate the role of human A20 variants in vivo. Scientific reports. 2020;10(1):19085.
403 4. Zammit NW, Walters SN, Seeberger KL, O'Connell PJ, et al. A20 as an immune tolerance
404 factor can determine islet transplant outcomes. JCI Insight. 2019;4(21).
405 5. Wertz IE, O'Rourke KM, Zhou H, Eby M, et al. De-ubiquitination and ubiquitin ligase domains
406 of A20 downregulate NF-kappaB signalling. Nature. 2004;430(7000):694-9.
407 6. Evans PC, Ovaa H, Hamon M, Kilshaw PJ, et al. Zinc-finger protein A20, a regulator of
408 inflammation and cell survival, has de-ubiquitinating activity. Biochem J. 2004;378(Pt 3):727-34.
409 7. Lu TT, Onizawa M, Hammer GE, Turer EE, et al. Dimerization and ubiquitin mediated
410 recruitment of a20, a complex deubiquitinating enzyme. Immunity. 2013;38(5):896-905.
411 8. Bosanac I, Wertz IE, Pan B, Yu C, et al. Ubiquitin binding to A20 ZnF4 is required for
412 modulation of NF-kappaB signaling. Mol Cell. 2010;40(4):548-57.
413 9. Wertz IE, Newton K, Seshasayee D, Kusam S, et al. Phosphorylation and linear ubiquitin
414 direct A20 inhibition of inflammation. Nature. 2015;528(7582):370-5.
415 10. Verhelst K, Carpentier I, Kreike M, Meloni L, et al. A20 inhibits LUBAC-mediated NF-kappaB
416 activation by binding linear polyubiquitin chains via its zinc finger 7. EMBO J. 2012;31(19):3845-55.
417 11. Liuwantara D, Elliot M, Smith MW, Yam AO, et al. Nuclear factor-kappaB regulates beta-cell
418 death: a critical role for A20 in beta-cell protection. Diabetes. 2006;55(9):2491-501.
419 12. Dixit VM, Green S, Sarma V, Holzman LB, et al. Tumor necrosis factor-alpha induction of
420 novel gene products in human endothelial cells including a macrophage-specific chemotaxin. Journal
421 of Biological Chemistry. 1990;265(5):2973-8.
422 13. Krikos A, Laherty CD, Dixit VM. Transcriptional activation of the tumor necrosis factor alpha-
423 inducible zinc finger protein, A20, is mediated by kappa B elements. J Biol Chem. 1992;267:17971-6.
424 14. Hutti JE, Turk BE, Asara JM, Ma A, et al. IkappaB kinase beta phosphorylates the K63
425 deubiquitinase A20 to cause feedback inhibition of the NF-kappaB pathway. Mol Cell Biol.
426 2007;27(21):7451-61.
427 15. Zammit NW, Siggs OM, Gray PE, Horikawa K, et al. Denisovan, modern human and mouse
428 TNFAIP3 alleles tune A20 phosphorylation and immunity. Nat Immunol. 2019;20(10):1299-310.
429 16. Ma A, Malynn BA. A20: linking a complex regulator of ubiquitylation to immunity and human
430 disease. Nat Rev Immunol. 2012;12(11):774-85.
431 17. Zhou Q, Wang H, Schwartz DM, Stoffels M, et al. Loss-of-function mutations in TNFAIP3
432 leading to A20 haploinsufficiency cause an early-onset autoinflammatory disease. Nat Genet.
433 2016;48(1):67-73.
434 18. Duncan CJA, Dinnigan E, Theobald R, Grainger A, et al. Early-onset autoimmune disease due
435 to a heterozygous loss-of-function mutation in TNFAIP3 (A20). Ann Rheum Dis. 2017.
436 19. Kadowaki T, Ohnishi H, Kawamoto N, Hori T, et al. Haploinsufficiency of A20 causes
437 autoinflammatory and autoimmune disorders. J Allergy Clin Immunol. 2017.
438 20. Ohnishi H, Kawamoto N, Seishima M, Ohara O, et al. A Japanese family case with juvenile
439 onset Behcet's disease caused by TNFAIP3 mutation. Allergol Int. 2017;66(1):146-8.
440 21. Shigemura T, Kaneko N, Kobayashi N, Kobayashi K, et al. Novel heterozygous C243Y
441 A20/TNFAIP3 gene mutation is responsible for chronic inflammation in autosomal-dominant
442 Behcet's disease. RMD Open. 2016;2(1):e000223.
443 22. Takagi M, Ogata S, Ueno H, Yoshida K, et al. Haploinsufficiency of TNFAIP3 (A20) by germline
444 mutation is involved in autoimmune lymphoproliferative syndrome. J Allergy Clin Immunol.
445 2017;139(6):1914-22.

446 23. Aeschlimann FA, Batu ED, Canna SW, Go E, et al. A20 haploinsufficiency (HA20): clinical
447 phenotypes and disease course of patients with a newly recognised NF-kB-mediated
448 autoinflammatory disease. Ann Rheum Dis. 2018;77(5):728-35.
449 24. Berteau F, Rouviere B, Nau A, Le Berre R, et al. 'A20 haploinsufficiency (HA20): clinical
450 phenotypes and disease course of patients with a newly recognised NF-kB-mediated
451 autoinflammatory disease'. Ann Rheum Dis. 2019;78(5):e35.
452 25. He T, Huang Y, Luo Y, Xia Y, et al. Haploinsufficiency of A20 Due to Novel Mutations in
453 TNFAIP3. J Clin Immunol. 2020;40(5):741-51.
454 26. Guggino WB, Stanton BA. New insights into cystic fibrosis: molecular switches that regulate
455 CFTR. Nature reviews Molecular cell biology. 2006;7(6):426-36.
456 27. Wexler NS, Lorimer J, Porter J, Gomez F, et al. Venezuelan kindreds reveal that genetic and
457 environmental factors modulate Huntington's disease age of onset. Proc Natl Acad Sci U S A.
458 2004;101(10):3498-503.
459 28. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, et al. The mutational constraint spectrum
460 quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
461 29. Gittelman RM, Schraiber JG, Vernot B, Mikacenic C, et al. Archaic Hominin Admixture
462 Facilitated Adaptation to Out-of-Africa Environments. Curr Biol. 2016;26(24):3375-82.
463 30. Karin M, Ben-Neriah Y. Phosphorylation meets ubiquitination: the control of NF-[kappa]B
464 activity. Annu Rev Immunol. 2000;18:621-63.
465 31. Mathes E, O'Dea EL, Hoffmann A, Ghosh G. NF-kappaB dictates the degradation pathway of
466 IkappaBalpha. EMBO J. 2008;27(9):1357-67.
467 32. Aeschlimann FA, Laxer RM. Haploinsufficiency of A20 and other paediatric inflammatory
468 disorders with mucosal involvement. Curr Opin Rheumatol. 2018;30(5):506-13.
469 33. Turer EE, Tavares RM, Mortier E, Hitotsumatsu O, et al. Homeostatic MyD88-dependent
470 signals cause lethal inflammation in the absence of A20. J Exp Med. 2008;205(2):451-64.
471 34. Beura LK, Hamilton SE, Bi K, Schenkel JM, et al. Normalizing the environment recapitulates
472 adult human immune traits in laboratory mice. Nature. 2016;532(7600):512-6.
473 35. Rosshart SP, Herz J, Vassallo BG, Hunter A, et al. Laboratory mice born to wild mice have
474 natural microbiota and model human immune responses. Science. 2019;365(6452).
475 36. Albers CA, Paul DS, Schulze H, Freson K, et al. Compound inheritance of a low-frequency
476 regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR
477 syndrome. Nat Genet. 2012;44(4):435-9, S1-2.
478 37. Hebbel RP. Sickle hemoglobin instability: a mechanism for malarial protection. Redox Rep.
479 2003;8(5):238-40.
480 38. Phornphutkul C, Introne WJ, Perry MB, Bernardini I, et al. Natural history of alkaptonuria. N
481 Engl J Med. 2002;347(26):2111-21.
482 39. Blau N, van Spronsen FJ, Levy HL. Phenylketonuria. Lancet. 2010;376(9750):1417-27.
483 40. Hakansson A, Tormo-Badia N, Baridi A, Xu J, et al. Immunological alteration and changes of
484 gut microbiota after dextran sulfate sodium (DSS) administration in mice. Clin Exp Med.
485 2015;15(1):107-20.
486 41. Johansson ME, Gustafsson JK, Sjoberg KE, Petersson J, et al. Bacteria penetrate the inner
487 mucus layer before inflammation in the dextran sulfate colitis model. PLoS One. 2010;5(8):e12238.
488 42. Johansson ME, Gustafsson JK, Holmen-Larsson J, Jabbar KS, et al. Bacteria penetrate the
489 normally impenetrable inner colon mucus layer in both murine colitis models and patients with
490 ulcerative colitis. Gut. 2014;63(2):281-91.
491 43. Vereecke L, Sze M, Mc Guire C, Rogiers B, et al. Enterocyte-specific A20 deficiency sensitizes
492 to tumor necrosis factor-induced toxicity and experimental colitis. J Exp Med. 2010;207(7):1513-23.
493 44. Scher JU, Sczesnak A, Longman RS, Segata N, et al. Expansion of intestinal Prevotella copri
494 correlates with enhanced susceptibility to arthritis. Elife. 2013;2:e01202.
495 45. Iljazovic A, Roy U, Galvez EJC, Lesker TR, et al. Perturbation of the gut microbiome by
496 Prevotella spp. enhances host susceptibility to mucosal inflammation. Mucosal Immunol. 2020.

46. Atarashi K, Tanoue T, Ando M, Kamada N, et al. Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial Cells. Cell. 2015;163(2):367-80.
47. Larsen JM. The immune response to Prevotella bacteria in chronic inflammatory disease. Immunology. 2017;151(4):363-74.
48. Caruso R, Mathes T, Martens EC, Kamada N, et al. A specific gene-microbe interaction drives the development of Crohn's disease-like colitis in mice. Sci Immunol. 2019;4(34).
49. Kullberg MC, Ward JM, Gorelick PL, Caspar P, et al. Helicobacter hepaticus triggers colitis in specific-pathogen-free interleukin-10 (IL-10)-deficient mice through an IL-12- and gamma interferon-dependent mechanism. Infect Immun. 1998;66(11):5157-66.
50. Gobert AP, Sagrestani G, Delmas E, Wilson KT, et al. The human intestinal microbiota of constipated-predominant irritable bowel syndrome patients exhibits anti-inflammatory properties. Scientific reports. 2016;6:39399.
51. Zhai R, Xue X, Zhang L, Yang X, et al. Strain-Specific Anti-inflammatory Properties of Two Akkermansia muciniphila Strains on Chronic Colitis in Mice. Front Cell Infect Microbiol. 2019;9:239.
52. Vereecke L, Vieira-Silva S, Billiet T, van Es JH, et al. A20 controls intestinal homeostasis through cell-specific activities. Nature communications. 2014;5:5103.
53. Altin JA, Tian L, Liston A, Bertram EM, et al. Decreased T-cell receptor signaling through CARD11 differentially compromises forkhead box protein 3-positive regulatory versus T(H)2 effector cells to cause allergy. J Allergy Clin Immunol. 2011;127(5):1277-85 e5.
54. Ronin E, Lubrano di Ricco M, Vallion R, Divoux J, et al. The NF-kappaB RelA Transcription Factor Is Critical for Regulatory T Cell Activation and Stability. Frontiers in immunology. 2019;10:2487.
55. Oh H, Grinberg-Bleyer Y, Liao W, Maloney D, et al. An NF-kappaB Transcription-Factor-Dependent Lineage-Specific Transcriptional Program Promotes Regulatory T Cell Identity and Function. Immunity. 2017;47(3):450-65 e5.
56. Zhang Z, Zhang W, Guo J, Gu Q, et al. Activation and Functional Specialization of Regulatory T Cells Lead to the Generation of Foxp3 Instability. J Immunol. 2017;198(7):2612-25.
57. Mikami N, Kawakami R, Chen KY, Sugimoto A, et al. Epigenetic conversion of conventional T cells into regulatory T cells by CD28 signal deprivation. Proc Natl Acad Sci U S A. 2020;117(22):12258-68.
58. Salama AD, Chitnis T, Imitola J, Ansari MJ, et al. Critical role of the programmed death-1 (PD-1) pathway in regulation of experimental autoimmune encephalomyelitis. J Exp Med. 2003;198(1):71-8.
59. Free ME, Bunch DO, McGregor JA, Jones BE, et al. Patients with antineutrophil cytoplasmic antibody-associated vasculitis have defective Treg cell function exacerbated by the presence of a suppression-resistant effector cell population. Arthritis Rheum. 2013;65(7):1922-33.
60. Valencia X, Yarboro C, Illei G, Lipsky PE. Deficient CD4+CD25high T regulatory cell function in patients with active systemic lupus erythematosus. J Immunol. 2007;178(4):2579-88.
61. Ooi JD, Snelgrove SL, Engel DR, Hochheiser K, et al. Endogenous foxp3(+) T-regulatory cells suppress anti-glomerular basement membrane nephritis. Kidney Int. 2011;79(9):977-86.
62. Wang KW, Zhan X, McAlpine W, Zhang Z, et al. Enhanced susceptibility to chemically induced colitis caused by excessive endosomal TLR signaling in LRBA-deficient mice. Proc Natl Acad Sci U S A. 2019;116(23):11380-9.
63. Chen H, Li H, Liu Z. Interplay of intestinal microbiota and mucosal immunity in inflammatory bowel disease: a relationship of frenemies. Therap Adv Gastroenterol. 2020;13:1756284820935188.
64. Mevissen TE, Kulathu Y, Mulder MP, Geurink PP, et al. Molecular basis of Lys11-polyubiquitin specificity in the deubiquitinase Cezanne. Nature. 2016;538(7625):402-5.
65. Siggs OM, Russell A, Singh-Grewal D, Wong M, et al. Preponderance of CTLA4 Variation Associated With Autosomal Dominant Immune Dysregulation in the MYPPPY Motif. Frontiers in immunology. 2019;10:1544.

547 66. Gayevskiy V, Roscioli T, Dinger ME, Cowley MJ. Seave: a comprehensive web platform for
548 storing and interrogating human genomic variation. bioRxiv. 2018.
549 67. Minoche AE, Lundie B, Peters GB, Ohnesorg T, et al. ClinSV: Clinical grade structural and copy
550 number variant detection from whole genome sequencing data. medRxiv.
551 2020(2020.06.30.20143453).
552 68. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, et al. Reproducible, interactive, scalable and
553 extensible microbiome data science using QIIME 2. Nature biotechnology. 2019;37(8):852-7.
554 69. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
555 EMBnetjournal. 2011;17(1):10-2.
556 70. Rognes T, Flouri T, Nichols B, Quince C, et al. VSEARCH: a versatile open source tool for
557 metagenomics. PeerJ. 2016;4:e2584.
558 71. Amir A, McDonald D, Navas-Molina JA, Kopylova E, et al. Deblur Rapidly Resolves Single-
559 Nucleotide Community Sequence Patterns. mSystems. 2017;2(2).
560 72. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, et al. Optimizing taxonomic classification of
561 marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome.
562 2018;6(1):90.
563 73. Mandal S, Van Treuren W, White RA, Eggesbo M, et al. Analysis of composition of
564 microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis.
565 2015;26:27663.
566 74. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and
567 graphics of microbiome census data. PLoS One. 2013;8(4):e61217.
568 75. Hamady M, Lozupone C, Knight R. Fast UniFrac: facilitating high-throughput phylogenetic
569 analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. ISME J.
570 2010;4(1):17-27.
571 76. M.J. A. Permutational multivariate analysis of variance (PERMANOVA). In: N. Balakrishnan
572 TC, B. Everitt, W. Piegorsch, F. Ruggeri and J.L. Teugels, editor. Wiley StatsRef: Statistics References
573 Online2017.
574 77. Oksanen JF, Blanchet G, Friendly M, Kindt R, et al. Vegan: Community Ecology Package. R
575 package version 2.5-6. https://CRAN.R-project.org/package=vegan2019 [
576 78. Faith DP. Conservation evaluation and phylogenetic diversity. Biological Conservation.
577 1992;61(1):1-10.
578 79. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, et al. Picante: R tools for integrating
579 phylogenies and ecology. Bioinformatics. 2010;26(11):1463-4.
580 80. Wickham H., François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. R
581 package version 1.0.2. https://CRAN.R-project.org/package=dplyr2020 [
582 81. Erben U, Loddenkemper C, Doerfel K, Spieckermann S, et al. A guide to histomorphological
583 evaluation of intestinal inflammation in mouse models. Int J Clin Exp Pathol. 2014;7(8):4557-76.
584 82. Untergasser A, Cutcutache I, Koressaar T, Ye J, et al. Primer3--new capabilities and
585 interfaces. Nucleic Acids Res. 2012;40(15):e115.
586 83. Meyer M, Kircher M, Gansauge MT, Li H, et al. A high-coverage genome sequence from an
587 archaic Denisovan individual. Science. 2012;338(6104):222-6.

588

FUNDING:

N.W.Z was supported by an Australian Postgraduate Award and is an International Pancreas and Islet Transplant Association (IPITA) Derek Gray Fellow. The Clinical Immunogenomics Research Consortium of Australasia (CIRCA) was supported by the Jeffrey Modell Foundation, the John Brown Cook Foundation, and the Sydney Children's Hospital Network. N.W.Z was supported by an Australian Postgraduate Award and is supported by the International Pancreas and Islet Transplant Association (IPITA) Derek Gray Scholarship. S.G.T was a Principal Research Fellow (1042925) and is currently an Investigator (Leadership 3) of the NHMRC. C.C.G. was supported by the Bill and Patricia Ritchie Chair and by an NHMRC Senior Principal Research Fellowship. The research was supported by grants to C.C.G. from the NIH (AI52127, AI054523, AI100627) and NHMRC (1016953, 585490, 1081858), D.L.B from the NHMRC (1176351) and to S.T.G. from the NIH (DK076169) and NHMRC (1130222, 1140691). S.T.G. was supported by an NHMRC Senior Research Fellowship (1140691).

AUTHOR CONTRIBUTIONS:

CCG – project initiation, coordination, and data interpretation.

PEG, OMS, AR, JBZ, MEC, PB-A, JT, MW, MJC, VG, AEM, MED, CCG – Clinical recruitment, whole genome sequencing analysis of HA20 patients.

NWZ, JYY, SGT, AR, STG – Molecular and immune phenotyping of human PBMCs.

NWZ, DC, STG – molecular characterization of A20 variants.

NWZ, JW, SNW, DC, RB, DZ, DLB, STG – mouse avatar characterization, pathology and disease models.
NWZ, CB, TP, STG – Microbiome challenge (DSS model; cohousing model) and microbiome analysis.

NWZ, PEG, OMS, CCG, STG - analysed data and co-wrote the manuscript.

ACKNOWLEDGMENTS:

We thank the Biological Testing Facility at the Garvan Institute of Medical Research for animal care.

We thank the Clinical Immunogenomics Research Consortium of Australasia for collecting and co-ordinating human samples (CIRCA).
https://www.garvan.org.au/research/collaborative-programs/circa/about-circa.

COMPETING INTERESTS:

All authors declare they have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

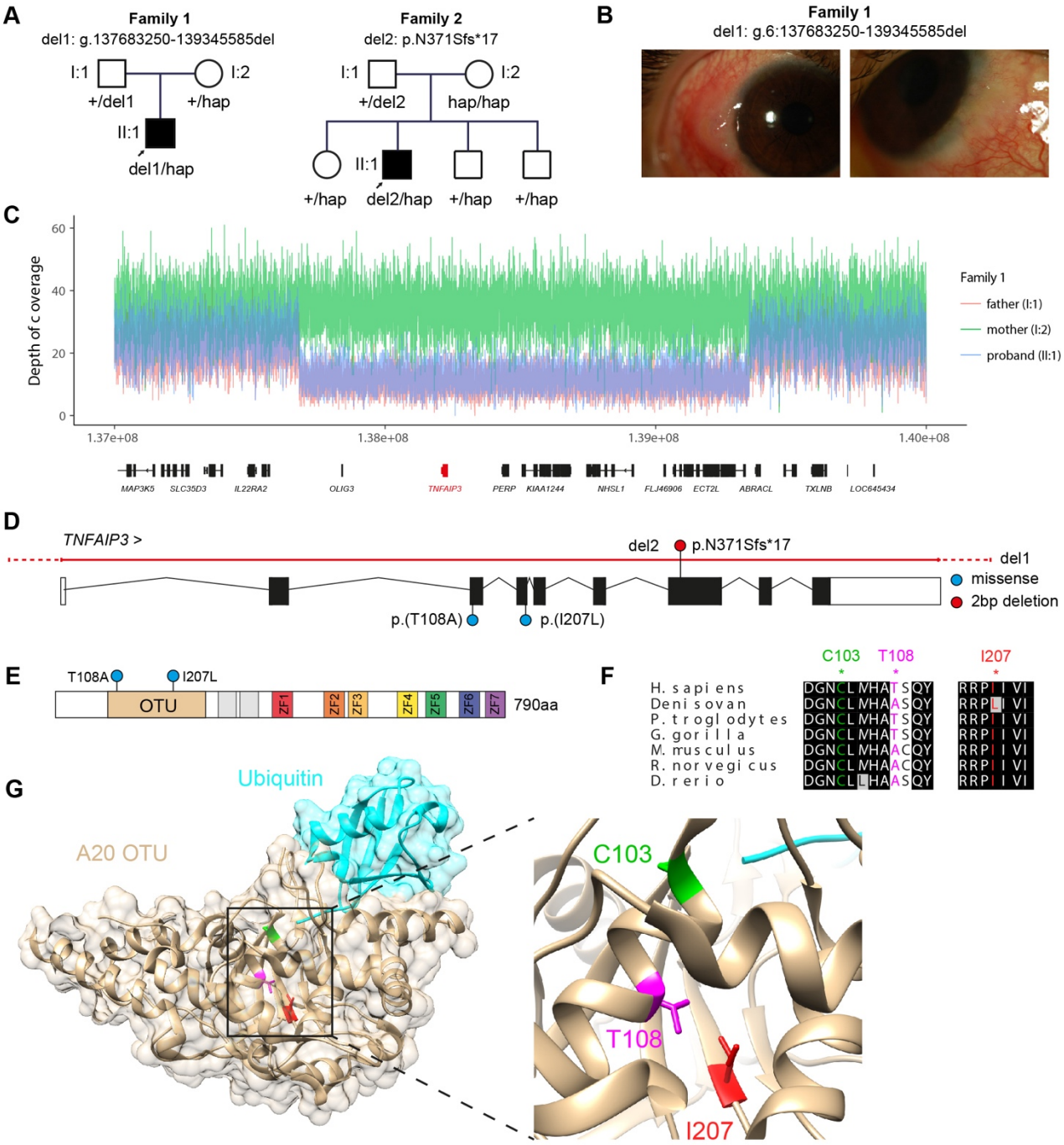
DATA AVAILABILITY:

The data that support the findings of this study are available from the corresponding author upon request

631 **FIGURES AND FIGURE LEGENDS:**

632

633



634

635 **Figure 1: HA20 probands harbour a coding denisovan introgressed *TNFAIP3* allele in**
636 ***trans* to A20 deletion.**
637 **(A)** Family trio's of identified probands. **(B)** Photo of ocular inflammation from proband of
638 family 1. **(C)** Sequencing coverage map depicting deletion of locus spanning *TNFAIP3. (***D)**
639 Schematic demonstrating the location of HA20 variants del1 from family 1 and del2 from
640 family 2. Each HA20 deletion is located *in trans* to two *in cis* missense variants T108A and
641 I207L, located in exons 3 and 4 of *TNFAIP3,* respectively. T108A and I207L formed part of a
642 larger haplotype adaptively introgressed from Denisovan Hominins (15, 29), referred to as hap
643 in A. **(E)** Location of *TNFAIP3* hap missense variants, T108A and I207L in reference to the
644 primary protein structure of A20. OTU = ovarian tumor domain. ZF = Zinc finger. **(F)**

645 Conservation and (**G)** location of T108 and I207 alleles in context with the OTU domain
646 catalytic 103 (5LRX)(64).

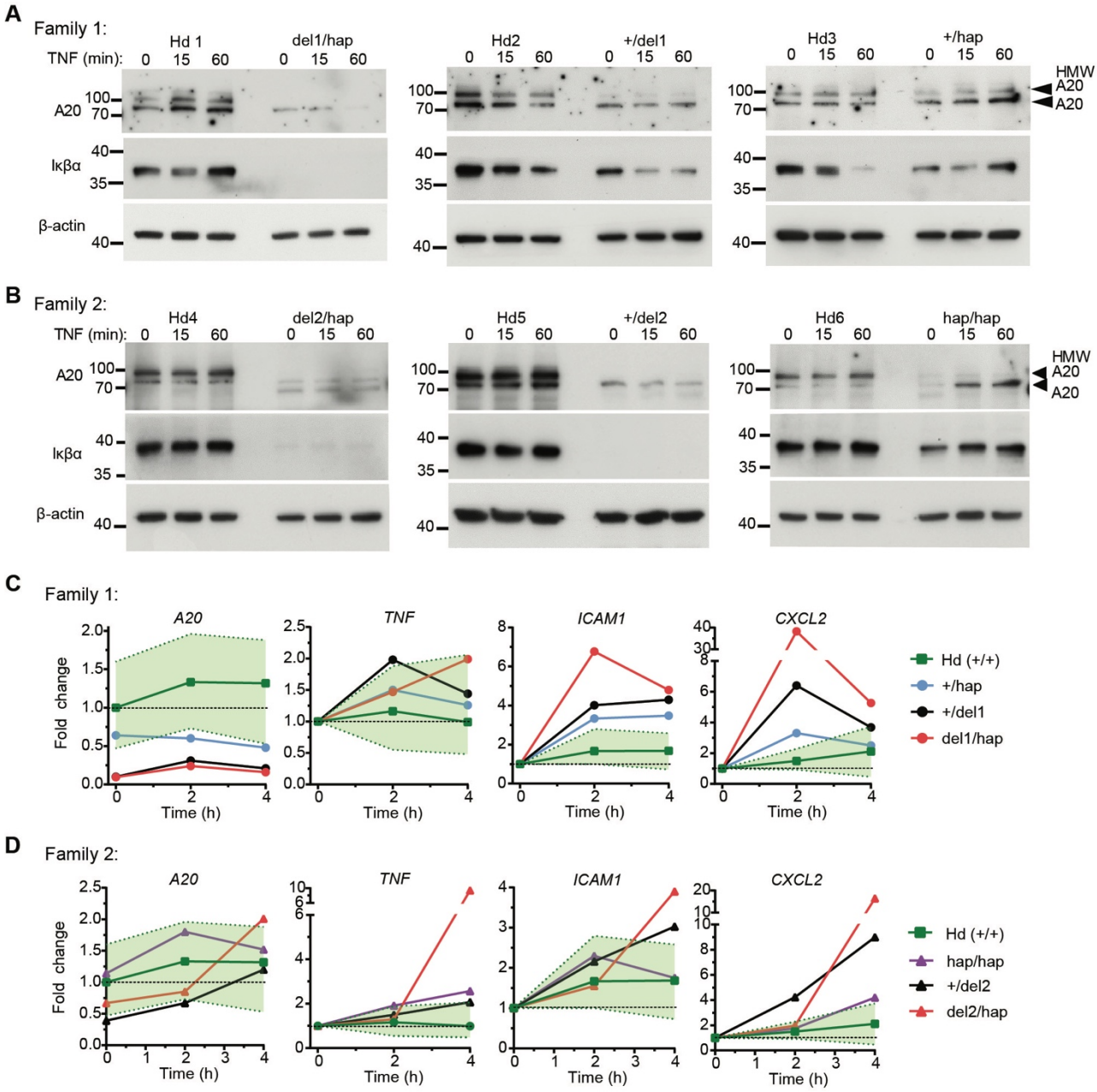


Figure 2: **Analysis of peripheral blood mononuclear cells (PBMCs) from parent-proband Trios.**

(A-B) Immunoblot analysis of PBMCs from parent-proband trios from family 1 **(A)** and family 2 **(B)**. Probands from each family harbor a *TNFAIP3* deletion (del1 or 2) and the *TNFAIP3* T108A;I207L oceanic haplotype (hap) in *trans* (del/hap). Parents are either heterozygous for a *TNFAIP3* del or heterozygous or homozygous for the *TNFAIP3* hap allele. PBMCs were left untreated or stimulated with TNF for 15 or 60 min. Proteins assessed included A20, IκBα and β-actin (loading control). A20 high-molecular weight bands (HMW) representing phospho-A20 are indicated (9, 14, 15). Healthy donors = Hd. Immunoblots of PBMC lysates from +/Hap and hap/hap parents (right panels) have been previously reported (15)(ref). **(C-D)**, RT–PCR analysis of *TNFAIP3*, *TNF, ICAM1* and *CXCL2* gene expression in PBMCs from proband trios in A and B. Probands (red), +/hap (blue), hap/hap (purple), +/del (black) and 15 healthy donors (green) +/- range (shaded light green).

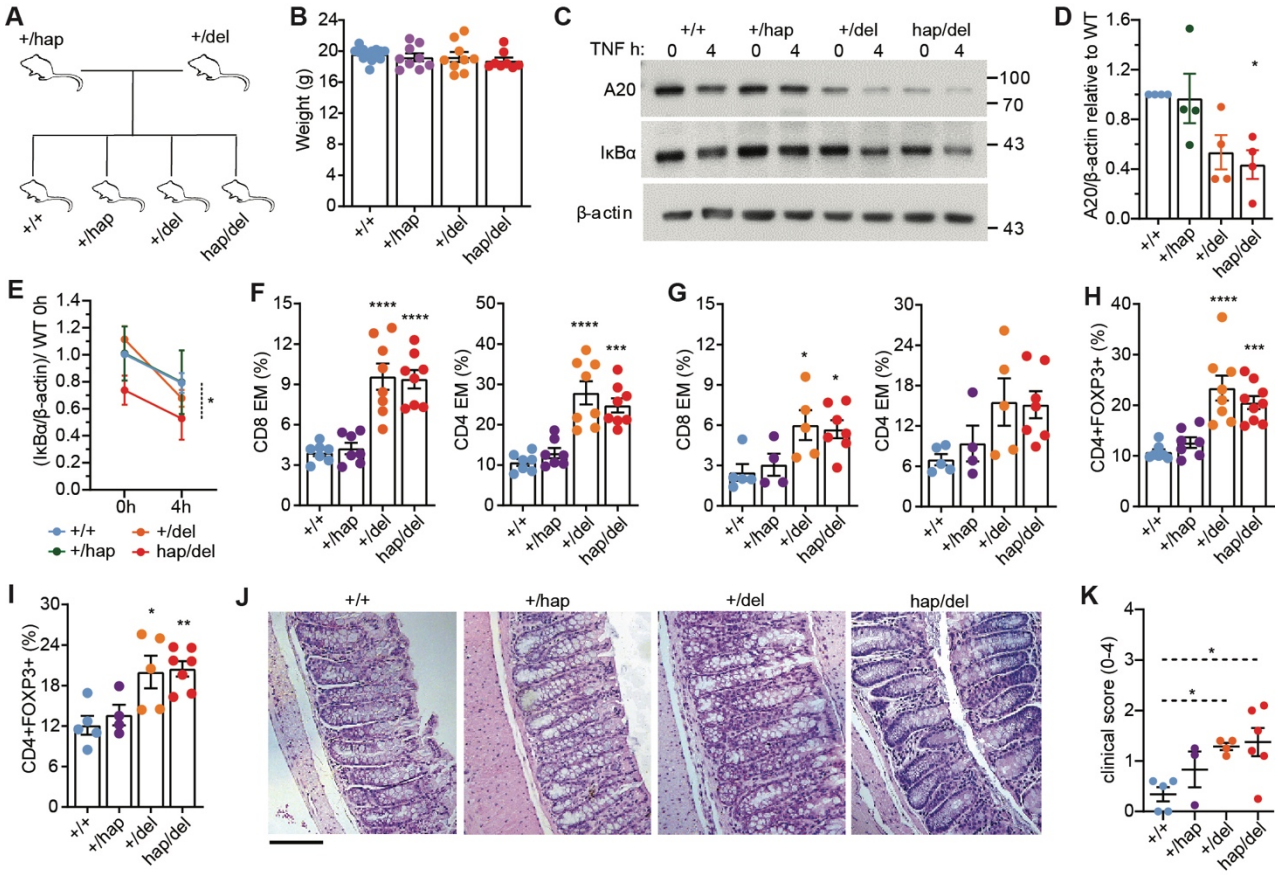


Figure 3: Molecular and phenotypical analysis of hap/del mice.
(A) Pedigree tree illustrating breeding strategy for *Tnfaip3* hap/del line. **(B)** Weight of 8-week-old female *Tnfaip3* hap/del mice of the indicated genotype. **(C)** Representative immunoblot of thymocyte lysates isolated from mice with indicated genotypes and stimulated with TNF for indicated times. Proteins probed include A20, IκBα and loading control β-actin. Cumulative densitometry is shown for **(D)** A20 and **(E)** IκBα with each point representing thymocytes from an individual mouse. **(F-G)** Percentage CD44hiCD62Lo effector memory cells among CD8+ or CD4+ T cells from the spleen (F) or Mesenteric lymph node (G) of mice with the indicated genotype. **(H-I)** Percentage of FOXP3+ regulatory T cells among CD4+ cells from the spleens (H) or mesenteric lymph nodes **(I)** of mice of the indicated genotypes. Flow cytometric data are cumulative from 3 independent experiments. **(J)** Representative sections showing H&E-stained colons from female mice of the indicated genotype. Scale bar = 200 µm. **(K)** Clinical score of representative colons shown in (J). Statistical analysis was performed by one-way anova ANOVA with Tukey post hoc test, and data shown as the mean +/- s.e.m: *P<0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

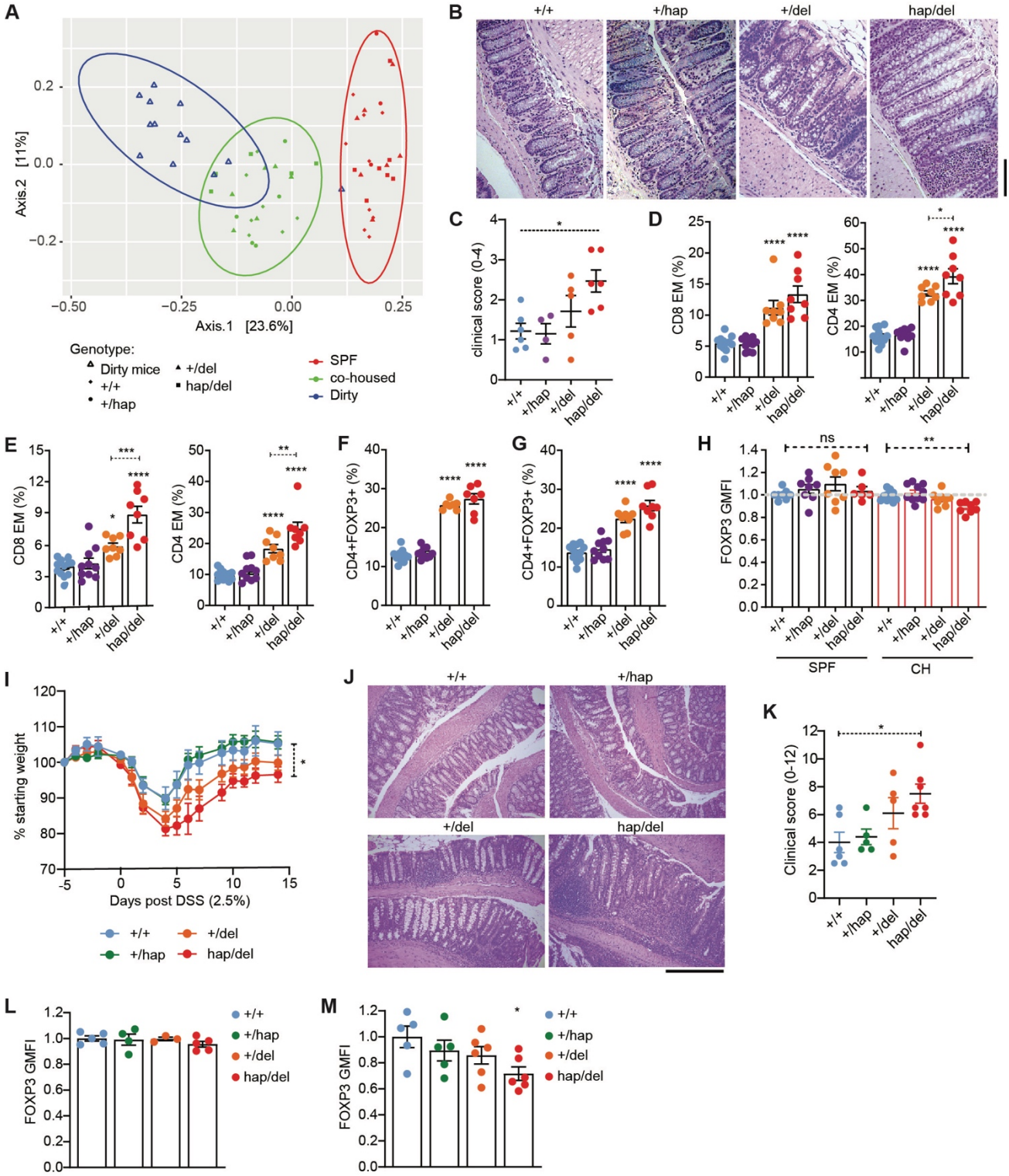


Figure 4: Microbiome and DSS challenge uncovers differential effects for hap/del mice compared to +/del alone.

(A) Unweighted unifrac PCoA showing cohoused *Tnfaip3* hap/del mice take on microbiota from dirty mice (P<0.001, PERMANOVA dirty mice vs SPF vs co-housed). Genotypes indicated by symbols. **(B)** Representative sections showing H&E-stained colons from female *Tnfaip3* hap/del mice 8 weeks post cohousing with dirty mice (20x magnification; scale bar = 200 µm), and **(C)** cumulative clinical score. **(D-E)** Percentage of CD44hiCD62Lo effector memory CD8+ or CD4+ T cells from the spleen (**D**) or mesenteric lymph node (**E**). **(F-G)**

688 Percentage of CD4+FOXP3+ regulatory T cells from the spleens (**F**) or mesenteric lymph
689 nodes (**G**) of mice of the indicated genotypes. **(H)** FOXP3 GMFI for CD4+FOXP3+ splenic
690 Tregs without cohousing (black bars) and post 8 weeks cohousing (red bars). FOXP3 GMFI
691 normalised to the average GMFI of +/+ samples **(I)** Percent change in starting weights of mice
692 with indicated genotypes given 2.5% DSS in drinking water for 5 days. Data is cumulative of
693 two independent experiments, +/+ n=6; +/hap n=5; +/del n=8; hap/del n=8. Statistical
694 significance determined by area under the curve analysis. **(J)** Representative haematoxylin and
695 eosin-stained colon sections 14 days post DSS treatment and **(K)** cumulative clinical scores.
696 **(L)** FOXP3 GMFI for CD4+FOXP3+ splenic Tregs under SPF conditions and **(M)** 14 days
697 post DSS treatment. FOXP3 GMFI normalized to the average GMFI of +/+ samples. Statistical
698 analysis was performed by one-way ANOVA with Tukey post hoc test unless otherwise stated,
699 and data shown as the mean +/- s.e.m: *$P < 0.05$; ****$P < 0.0001$. ns = not significant.

700 **Methods:**
701
702 Human subjects
703 Patients and healthy family members were recruited via the Clinical Immunogenomics
704 Research Consortium Australia (CIRCA) and The Children's Hospital at Westmead, providing
705 written informed consent under protocols approved by the Sydney *Children's Hospitals*
706 Network Human Research *Ethics* Committee (SCHN HREC). Blood samples from healthy
707 volunteers were obtained from the Australian Red Cross Blood Service.
708
709 Genome sequencing
710 Parent-proband trio genomes were sequenced on the Illumina HiSeq X platform using
711 DNA isolated from whole blood, to 30x genome-wide coverage. Libraries were generated
712 using either Illumina TrueSeq PCR-free or TrueSeq Nano. Alignment, variant calling, and
713 annotation was performed as described previously (65-67).
714
715 Mice
716 To generate *Tnfaip3*^{246tr} and *Tnfaip3*^{I207L} strains, respective guide RNAs (5′-
717 GGGATATCTGTAACACTCC -3′ and 5′-TGACAATGATGGGTCTTCTGAGG-3′) were
718 microinjected into C57BL/6 zygotes in combination with Cas9 mRNA. For *Tnfaip3*^{I207L} strain
719 an oligo template was also co-injected
720 (5′CTCCTCAGAGCTGAAACTCACCCAGGGAACCTAGAAACTCTCTGAGGCACCTC
721 ACCTGAAATGACAATGAGGGGTCTTCTGAGAATGTTGCTGAGGACAAATATGTG
722 GATTTCTTCCAGGGAATTGTACTGAAGTCCACTTCGGGCTGCA-3′). Founder mice
723 carrying the appropriate substitutions were then crossed to C57BL/6, and heterozygous
724 offspring intercrossed to generate homozygous, heterozygous and wild-type littermates. Mice
725 were housed at the Australian Phenomics Facility (Australian National University, Canberra,
726 Australia), or at the Australian BioResources Centre (ABR) (Moss Vale, NSW, Australia).
727 Animal studies were approved by the Garvan/St Vincent's or the Australian National
728 University Animal Ethics Committees. All procedures performed complied with the Australian
729 Code of Practice for Care and Use of Animals for Scientific Purposes.
730
731 Microbiome 16S rRNA sequencing and analysis
732 Fecal pellets from individual mice were stored at -80 prior to processing. DNA was extracted
733 using the PowerFecal DNA extraction kit (MoBio), including blanks as negative controls. 16S
734 rRNA gene libraries were prepared by amplifying the V4 region with primers that contained
735 the 515F (GTGYCAGCMGCCGCGGTAA) and 806R (GGACTACNVGGGTWTCTAAT)
736 sequences, along with 0-3 bp phasers and partial Truseq Illumina adaptors, and KAPA HiFi
737 Hotstart ready mix for 20 amplification cycles. PCR products were diluted 1:200 and amplified
738 for a further 10 cycles with primers that included the TruSeq Illumina adaptors and 6bp sample
739 barcodes on each primer. Both DNA extraction blanks and PCR negative controls were
740 included. Final PCR products were cleaned with 0.8V JetSeq Clean Beads (Bioline) and PCR
741 yield measured with a Quant-it PicoGreen ds DNA assay (ThermoFisher). Samples with a
742 detectable PCR yield were pooled at an equimolar concentration, while negative controls or
743 samples that had no detectable yield were added to the pool at the median volume used for all
744 other samples. The pool was quantified with a Bioanalyser HS DNA chip. The library was
745 sequenced on an Illumina MiSeq using a V2 flow cell with 2 x 250 bp paired end reads.
746
747 Sequencing data was assessed for quality using FastQC, and processed using Qiime2 v 2019.10
748 (68). In brief, primers and spacers were trimmed using cutadapt (69), forward and reverse
749 reads were merged using vsearch (70) and sequences were quality filtered and denoised to

750 amplicon sequence variants (ASVs) using Deblur (71). ASVs were classified using a naive
751 Bayes machine-learning classifier (72) trained on the V4 region of the Greengenes 16S rRNA
752 gene operational taxonomic unit (OTU) database clustered at 99% (accessed from
753 https://docs.qiime2.org/2019.10/data-resources/ in November 2019). A rooted phylogenetic
754 tree was constructed using the phylogeny plu-in and align-to-tree-mafft-fasttree
755 command. Differential relative abundance analysis was performed using the composition
756 plug-in and ANCOM (73).
757
758 Further analysis was performed in the R environment. Phyloseq (74) was used to import and
759 transform ASV counts and sample metadata. Samples with less than 10x the largest sequence
760 count of any negative controls were removed from the analysis. ASV counts were log
761 transformed prior to calculating weighted and unweighted unifrac distances using a Phyloseq
762 wrapper of FastUnifrac (75), and used as input for Principal Coordinates Analysis in
763 Phyloseq. Significant differences in clustering were assessed with PERMANOVA (76) via the
764 adonis function of Vegan (77). Faith's phylogenetic diversity (78) was calculated with Picante
765 (79) and differences in diversity between groups assessed with ANOVA in base R. Data
766 frames were manipulated using dplyr (80) and plots generated with ggplot2 (80).
767
768 Immunohistochemistry
769 Tissues were fixed in 10% neutral buffered formalin (Sigma-Aldrich), paraffin embedded
770 and sections (5 µm) prepared and stained with hematoxylin and eosin (H&E; Sigma-Aldrich).
771 For clinical scoring 4-6 random field of views at 10x magnification from three sections 500
772 µm apart were assessed per colon. Scoring of colons from non-DSS treated mice was carried
773 out as described by Erben et al 2014 (81), Scoring scheme 2 for colonic inflammation mediated
774 by luminal antigens. Briefly, colon sections were scored (1-4) according to two criteria, a) the
775 severity of inflammatory cell infiltrate and, b) epithelial changes including hyperplasia, goblet
776 cell and crypt loss. The scores for each criteria were averaged to give the final score, with 4
777 representing the highest score. Scoring of colons from DSS treated mice was carried out as
778 described by Erben et al 2014 (81). Briefly, colon sections were scored on four criteria, 1,
779 Inflammatory cell infiltrate, 2, erosion and ulceration, 3, submucosal edema and hyperplasia
780 and 4, loss of crypts and goblet cells, for a total clinical score of 12. Images were captured
781 using a Leica DM 4000 or Leica DM 6000 Power Mosaic microscope (Leica Microsystems).
782
783 Flow cytometry
784 Flow cytometric staining was performed as described (15). Fluorochrome-conjugated
785 antibody clones against the following surface antigens were: CD3 (17A2), CD45.2 (104), B220
786 (RA3-6B2), IgD (11-26), Ly6G (IA8) (Biolegend). CD4 (RM4-5), CD8 (53-6.7), CD11b
787 (M1/70), CD23 (B3B4), CD44 (IM7), FOXP3 (FJK-165) (Ebioscience). CD21 (7G6), CD25
788 (PC61.5), CD62L (MEL-14) CD138 (281-2), FAS (Jo2), IgM (II/41) (BD Biosciences). Data
789 were acquired with FACSCanto II and FORTESSA flow cytometers (BD Biosciences) and
790 analyzed using FlowJo software (Tree Star).
791
792 Immunoblot analysis and immunoprecipitation
793 PBMC's and murine thymocytes were lysed using lysis buffer (Cell Signaling
794 Technology). Protein concentration was measured using the Bradford assay (Bio-Rad) and total
795 protein (20-25 µg) resolved on a 7 – 10% SDS PAGE gel and then transferred to a nitrocellulose
796 membrane, Immobilon-P® (Merck Millipore). Membranes were incubated with anti-IκBα
797 (product #: 9242), anti-A20 (Product #: 5630; clone: D13H3) (Cell Signaling Technology),
798 anti-A20 (product #: ab13597; clone: 59A426) (Abcam) or anti-beta-actin (Product #: A1978;
799 clone: AC15) (Sigma-Aldrich); followed by horseradish peroxidase (HRP)-labeled secondary

antibody goat-anti-mouse IgG Fc (Product #: 31433) (Pierce Antibodies) or donkey-anti-rabbit IgG (Product #: NA934V) (GE Life Sciences). HRP conjugates bound to antigen were detected and visualized by using an ECL detection kit (GE Life Sciences).

Real Time quantitative PCR:

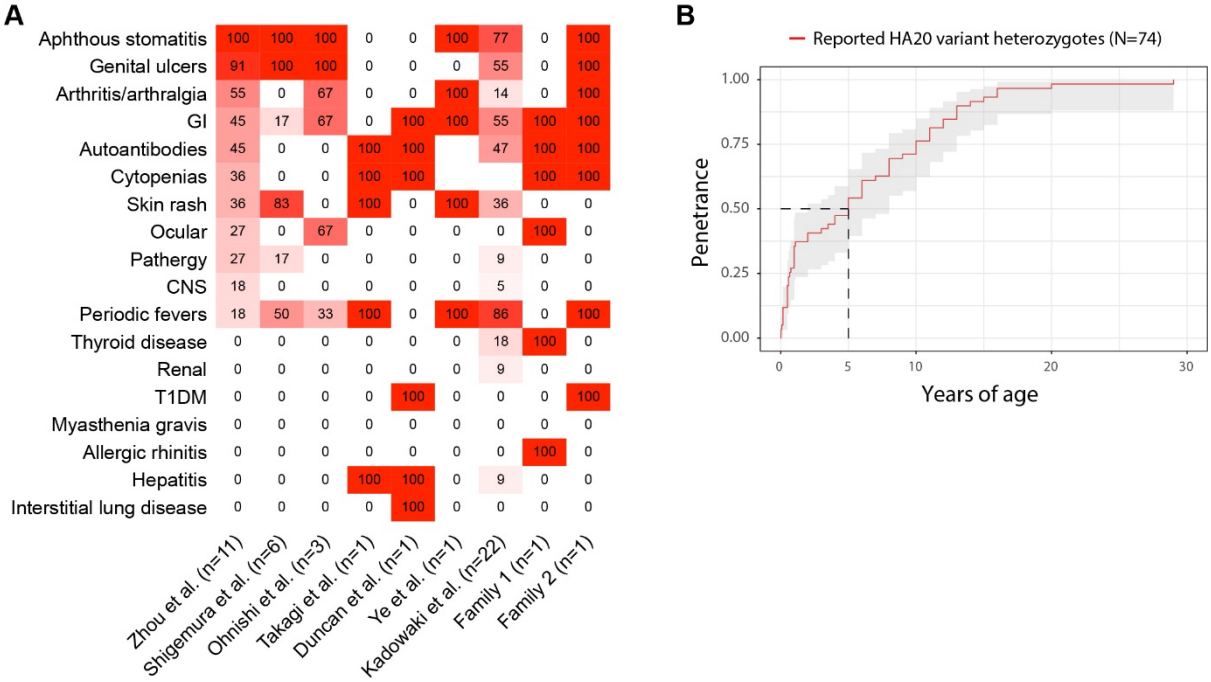
Total RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) and reverse transcribed using Quantitect Reverse Transcription Kit (Qiagen). Primers were designed using Primer3 (82) with sequences obtained from GenBank and synthesized by Sigma Aldrich (Supplementary Table 1). PCR reactions were performed on the LightCycler® 480 Real Time PCR System (Roche) using the FastStart SYBR Green Master Mix (Roche). *RPL13A* and ACTB were used as housekeeping genes and data analysed using the 2ΔΔCT method. Initial denaturation was =performed at 95° C for 10 sec, followed by a three-step cycle consisting of 95° C for 15 sec =(4.8° C/s, denaturation), 63° C for 30 sec (2.5° C/sec, annealing), and 72° C for 30 sec (4.8° C/s, elongation). A melt-curve was performed after finalization of 45 cycles at 95° C for 2 min, 40° C for 3 min and gradual increase to 95° C with 25 acquisitions/° C.

Statistical methods

Results are expressed as mean +/- standard error mean (SEM). Statistical analysis was performed using the Student's *t*-test or ANOVA where indicated.

819 **Supplementary Figures and legends:**

820

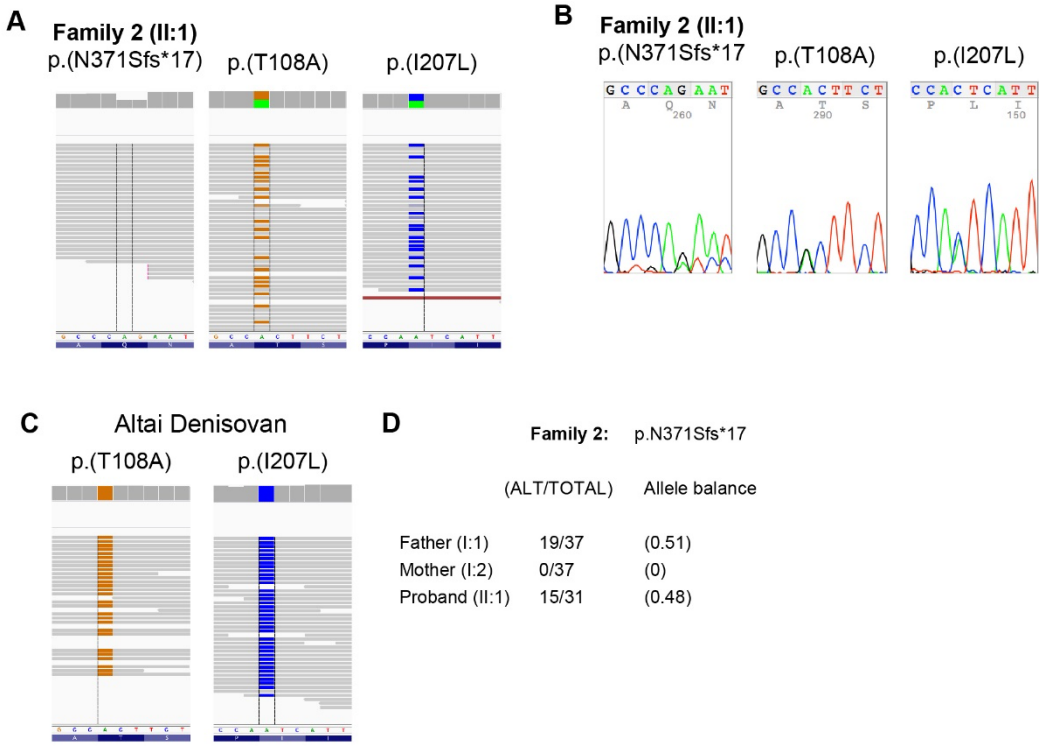


821

822 **Supplementary Figure 1**

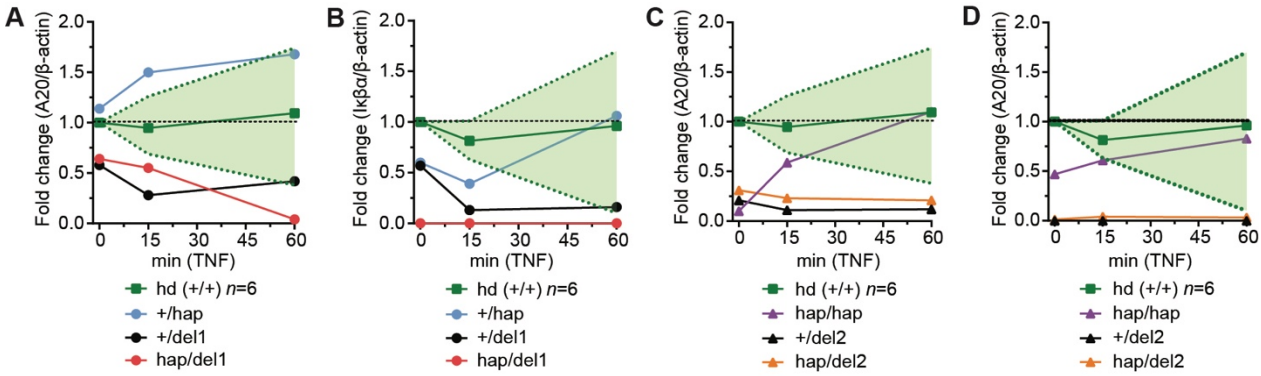
823 **(A)** Summary of clinical presentations observed in probands from family 1 and 2 in this
824 study, compared to reported cases (18-22). **(B)** Age of disease of 74 reported HA20 variant
825 heterozygotes.

826



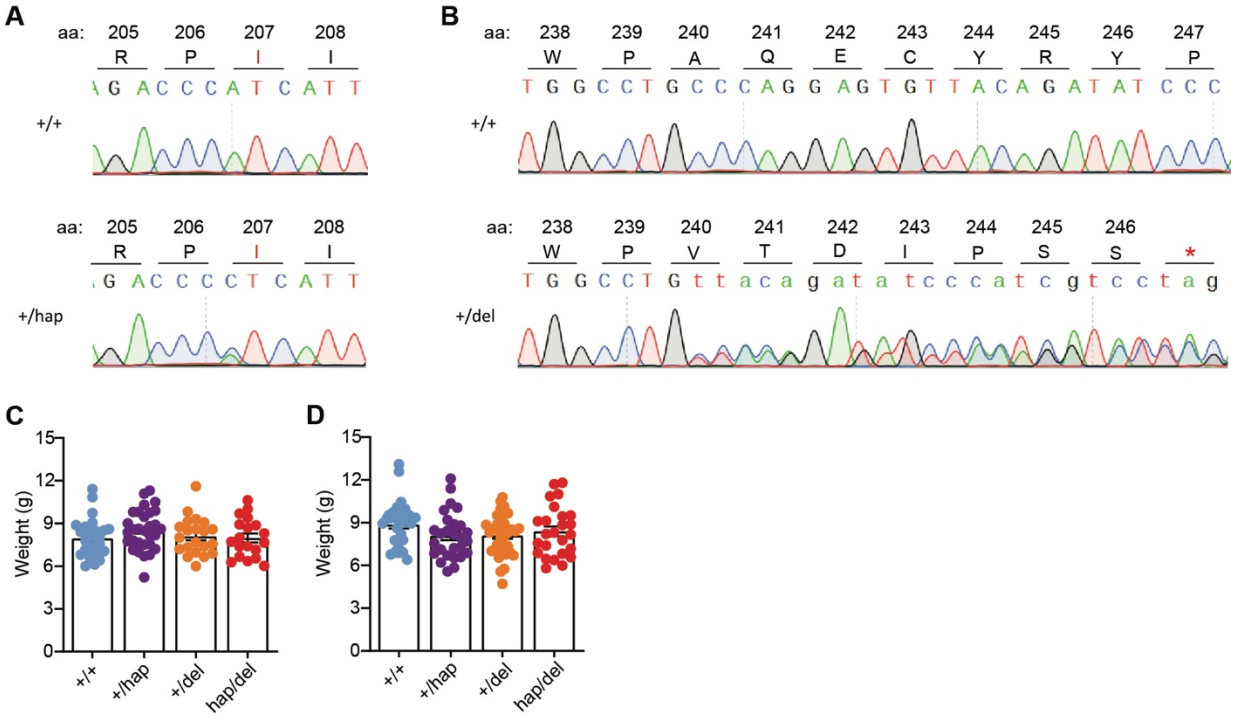
Supplementary Figure 2

(A) Read data of family 2 across TNFAIP3 frame shift (N371Sfs*17) and codols 108 and 207 **(B)** and representative chromatograms. **(C)** Read data of a high-coverage Denisovan genome (83) across TNFAIP3 codons 108 and 207. **(D)** Read call outs of p.N371Sfs*17 in family 2 trio.

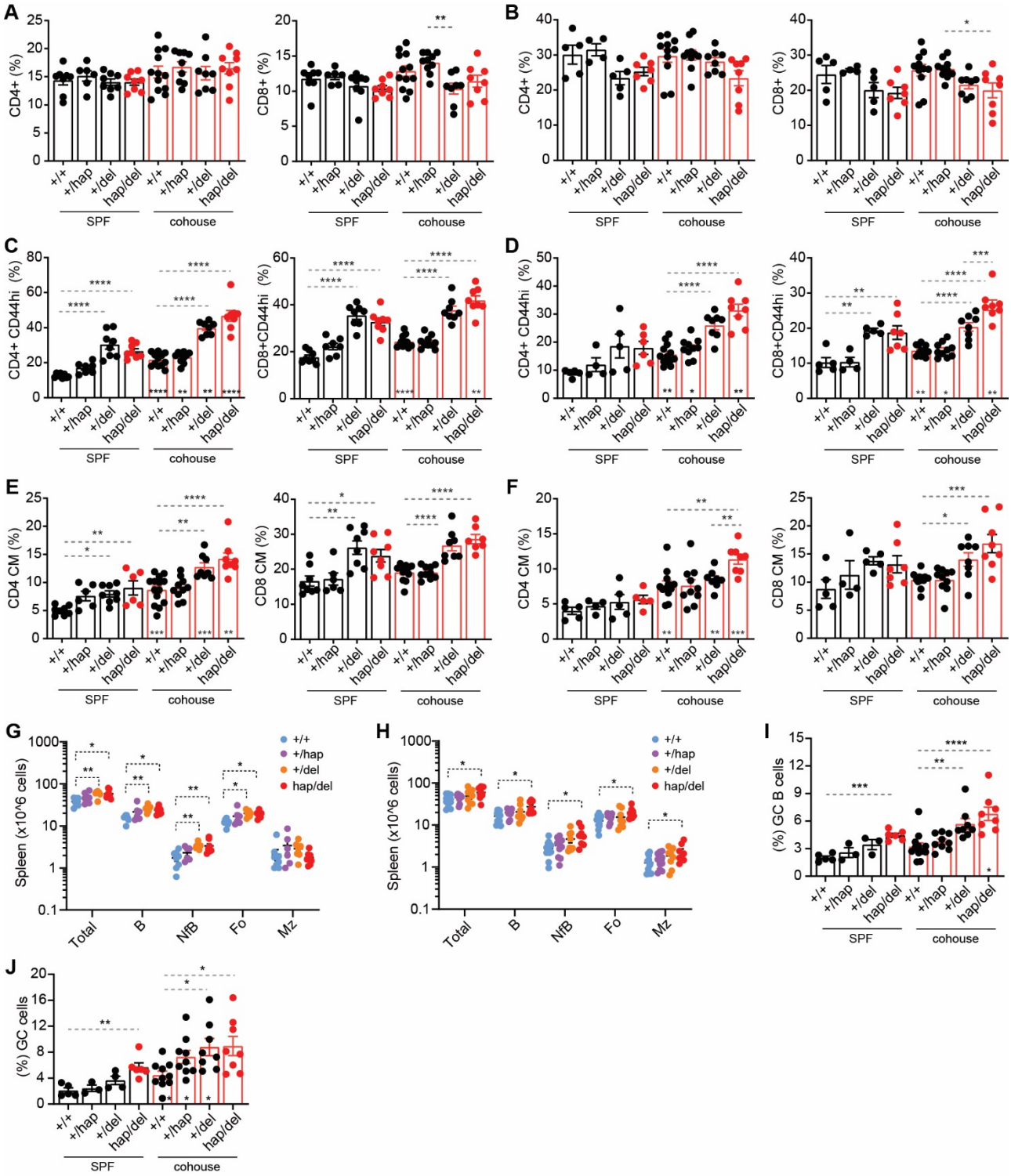


Supplementary Figure 3

(A-D) Densitometry analysis of A20 (A, C) or IκBα levels (B, D) from Trio 1 (A, B) and Trio 2 (C, D) immunoblots in Fig 2A, B. Six healthy donor (Hd) values presented as +/- range (light green shade).



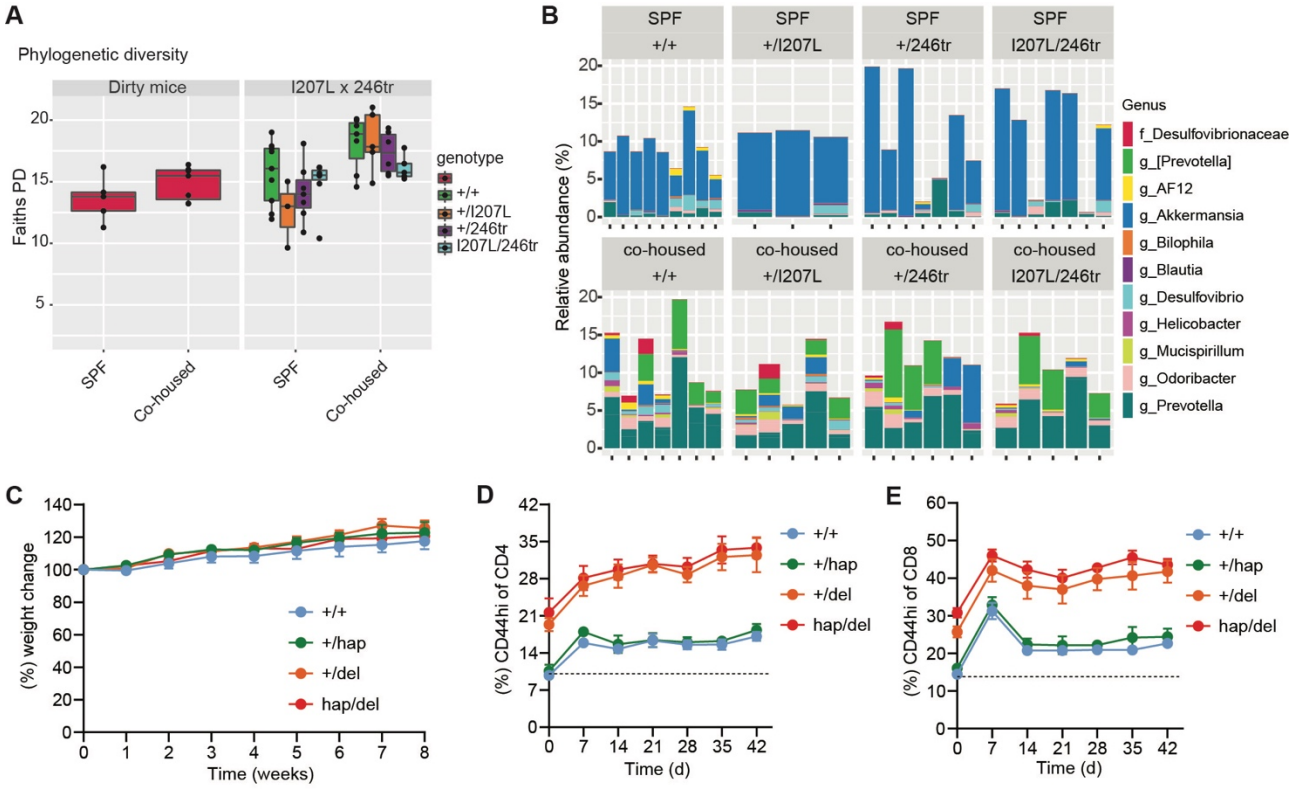
Supplementary Figure 4.
(A) Representative Sanger chromatogram for mice with the I207L (hap) or **(B)** A240Vfs*7 (del) alleles. **(C-D)** Weight of 8-week-old female (C) or male (D) *Tnfaip3* hap/del mice of the indicated genotype.



Supplementary Figure 5.

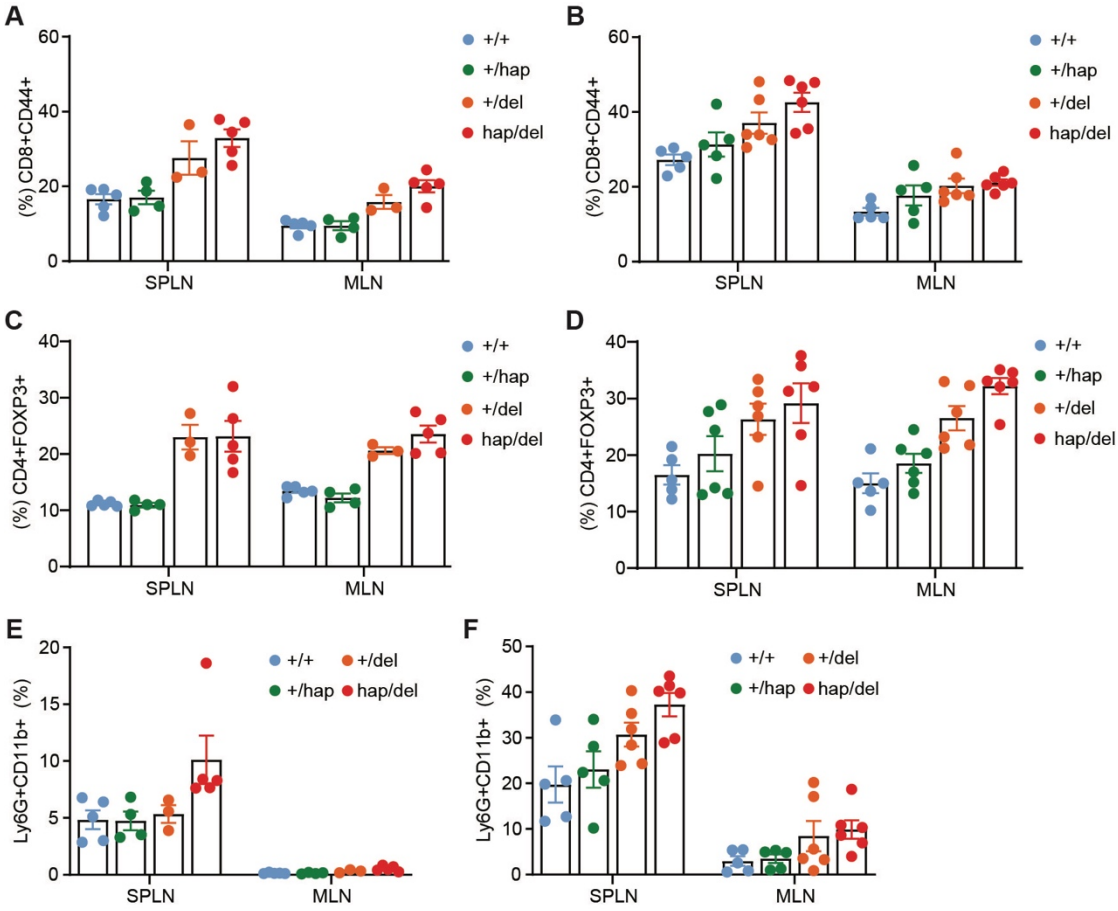
(A-F) Flow cytometry analysis of splenic (A, C, E) or mesenteric lymph node (B, D, F) T cell populations from Tnfaip3I207L/246tr mice housed under SPF conditions (black bars) or cohoused with dirty mice for 8 weeks (red bars). **(A, B)** Percentage of CD4+ or CD8+ T lymphocytes. **(C, D)** Percentage of CD4+ or CD8+ CD44hi memory T cells. **(E, F)** CD44hiCD62hi central memory CD8+ or CD4+ T cells. **(G, H)** Absolut numbers of B cells in the spleen of individual mice with the indicated *Tnfaip3* genotype, in SPF conditions (G) or post cohouse with dirty mice for 8 weeks (H). B, B cells; NFB, newly formed B cells; FO, follicular B cells; MZ, marginal zone B cells.

854 **(I-J)** Flow cytometry analysis of splenic (I) or mesenteric lymph node (J) B cell germinal
855 centres. **(K-M)** Percent proportions of splenic inflammatory monocytes (K), neutrophils (L)
856 and NK1.1 (M) positive cells were also assessed.**n**Statistical analysis was performed by one-
857 way anova ANOVA with Tukey post hoc test, and data shown as the mean +/- s.e.m: *P <
858 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001. ns = not significant.



Supplementary Figure 6.

(A) Bar plot showing relative abundance of significantly changing bacteria species following cohousing ($p<0.05$, two-way ANOVA). **(B)** Weekly weights presented as percent change in starting weight, for *Tnfaip3* hap/del mice of the indicated genotypes cohoused with dirty mice (+/+ n=4; +/hap n=5; +/del n=3; hap/del n=3). **(C)** Percent CD44hi CD8+ T cells or **(D)** CD44hi CD4+ T cells in the blood of *Tnfaip3* hap/del mice co-housed with dirty mice (+/+ n=7; +/hap n=6; +/del n=5; hap/del n=5).



Supplementary Figure 7.
(A, B) Percentage of CD44hi CD8+ or CD4+ T cells from the spleen (SPLN) or Mesenteric lymph node (MLN) of *Tnfaip3* hap/del mice of the indicated genotype (B) before or (C) 14 days post DSS treatment. **(C, D)** Percentage of CD4+FOXP3+ cells from the spleen (SPLN) or Mesenteric lymph node (MLN) of *Tnfaip3* hap/del mice of the indicated genotype (D) before or (E) 14 days post DSS treatment. **(E, F)** Percentage of CD45+Ly6G+CD11b+ neutrophils from the spleen (SPLN) or Mesenteric lymph node (MLN) of *Tnfaip3* hap/del mice of the indicated genotype (F) before or (G) 14 days post DSS treatment.

876 **Supplementary Tables**

877 **Supplementary Table 1.**

878 Human primers used for qRT-PCR analysis.

Gene	Primer sequence
ACTB	F- 5'-CATCGAGCACGGCATCGTCA-3′ R- 5′-TAGCACAGCCTGGATAGCAAC-3′.
CXCL2	F- 5'-AGGGAATTCACCTCAAGAACATCC-3' R- 5'-TGTGGCTATGACTTCGGTTTGG-3'
ICAM1	F- 5'-AGTTGCTCCTGCCTGGGAAC-3' R- 5'-TTTAGCTGTTGACTGCCCATCAG-3'
RPL13A	F- 5'-CAAGCGGATGAACACCAAC-3' R- 5'-TGTGGGGCAGCATACCTC-3'
TNFAIP3	F- 5'-AACGAACGGTGACGGCAAT-3' R- 5'-GAAGTCCACTTCGGGCCAT-3'
TNF	F- 5'-TCCCCTGGAAAGGACACCAT-3' R- 5'GGGTTTGCTACAACATGGGCT-3'

879