LRRK2 but not ATG16L1 is associated with Paneth cell defect in Japanese Crohn’s disease patients

Ta-Chiang Liu,1 Takeo Naito,2 Zhenqiu Liu,3 Kelli L. VanDussen,1 Talin Haritunians,3 Dalin Li,3 Katsuya Endo,2 Yosuke Kawai,4 Masao Nagasaki,3 Yoshitaka Kinouchi,5 Dermot P.B. McGovern,3 Tooru Shimosigawa,7 Yoichi Kakuta,2 and Thaddeus S. Stappenbeck1

1Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, USA. 2Division of Gastroenterology, Tohoku University Graduate School of Medicine, Sendai, Japan. 3F. Widjaja Family Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. 4Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan. 5Health Administration Center, Center for the Advancement of Higher Education, Tohoku University, Sendai, Japan.

Introduction

Crohn’s disease (CD) and ulcerative colitis are 2 main classical types of inflammatory bowel disease (IBD) (1, 2). The etiology of IBD involves genetic susceptibility and environmental triggers. Over 200 single nucleotide polymorphisms (SNPs) have been associated with susceptibility to IBD (3–7). This complex genetic network indicates that IBD likely encompasses more than the 2 classical subtypes. Therefore, novel, rationally designed biomarkers that can lead to disease stratification and personalized treatments are needed (8). One candidate method to subtype CD is to define the morphological patterns of small intestinal Paneth cells based on the intracellular distribution of granules containing antimicrobial proteins (Paneth cell phenotypes) (7). Paneth cells are specialized secretory cells located at the bases of the crypts of Lieberkühn in the small intestine (9–11). These cells produce a wide repertoire of antimicrobial
peptides, such as lysozyme and α-defensins, to modulate the intestinal microbiome (12–15) and thus are important mediators of the host innate immune response (16, 17).

We have previously shown that in European ancestry CD patients from North America, abnormal Paneth cell phenotype is associated with variants in \( \text{ATG16L1} \) and \( \text{NOD2} \) genes, both of which are CD susceptibility genes involved in autophagy (18, 19). Furthermore, we have also demonstrated that abnormal Paneth cell phenotype is associated with mucosal dysbiosis and aggressive disease course (19, 20). Thus, Paneth cell phenotype is a biologically and clinically relevant biomarker that can stratify CD patients.

To broaden the applications of Paneth cell phenotype in CD, a more detailed understanding of the genetic determinants of Paneth cell phenotype and further clinical validation are critical. However, as over 90% of Western CD cases are composed of patients of European ancestry who harbor \( \text{ATG16L1} \) T300A and/or \( \text{NOD2} \) CD-risk variants (3, 4), identification of novel genetic determinants in this population is challenging. Thus, studying CD patients from other ethnicities, especially those who possess a spectrum of susceptibility genes distinct from those found in European ancestry CD, would be helpful in identifying novel genetic determinants of Paneth cell phenotype. The genetic landscape of CD patients from Korea and Japan is partially overlapping but largely different than that of European ancestry CD (5, 21–23). In particular, the \( \text{NOD2} \) variants associated with European ancestry CD are not polymorphic in Japanese populations (24). In addition, while \( \text{ATG16L1} \) T300A is polymorphic in Japanese cohorts, it has not been associated with CD susceptibility (25). In contrast, SNPs that are more frequently associated with CD in Asian populations include those that tag \( \text{TNFSF15} \), \( \text{IL23R} \), \( \text{ATG16L2} \), \( \text{STAT3} \), \( \text{GPR35} \), \( \text{MHC Class II} \), and \( \text{ZNF365} \), among others (5, 21–23, 25). Therefore, a Japanese CD patient cohort represents a unique cohort to identify novel genetic determinants of Paneth cell phenotypes.

In this study, we hypothesized that novel genetic determinants of Paneth cell defect exist in Japanese CD, and that abnormal Paneth cell phenotype would be associated with poor outcome in Japanese CD patients. We further hypothesized that these novel genetic determinants of Paneth cell defect in Japanese CD would also be associated with autophagy as in North American CD. We found that the prevalence of abnormal Paneth cell phenotype in Japanese CD was similar to North American CD. Abnormal Paneth cell phenotype also correlated with more aggressive outcome in Japanese CD. Our hypothesis-driven analysis identified that autophagy-associated \( \text{LRRK2} \), but not \( \text{ATG16L1} \), was associated with Paneth cell defect in Japanese CD. In contrast, in North American CD, \( \text{ATG16L1} \) but not \( \text{LRRK2} \) was associated with Paneth cell defect. In addition, unbiased genome-wide analysis for Paneth cell phenotype genetic associations identified additional candidate SNPs predicted to interact with known CD susceptibility genes and jointly affect autophagy or TNF-α signaling.

**Results**

Paneth cell defects were prevalent in Japanese CD patients. The overall work flow for the determination of genetic associations of Paneth cell defect in Japanese CD patients is outlined in Figure 1. We first determined the percentage of normal Paneth cells in resection specimens of adult Japanese CD subjects using a previously reported method of immunofluorescence localization for defensin-5 (20). The staining for defensin-5 was specific for Paneth cells and the results were highly correlative to that obtained by lysozyme immunofluorescence (\( R^2 = 0.6451; P < 0.0001 \)) (Supplemental Figure 1; supplemental material available online with this
There was a similar range of intracellular staining patterns to those previously observed in multiple North American CD cohorts (19, 20) (Figure 2A). The average percentage of normal Paneth cells was similar between Japanese and North American CD subjects (82.13 ± 1.06 vs. 81.9 ± 0.55; \( P = 0.87 \)) (Figure 2B). As in previous studies, we utilized a cutoff of greater than or equal to 20% (high abundance) abnormal Paneth cells to define a type I Paneth cell phenotype (type II is < 20% abnormal Paneth cells). We found that the prevalence of type I Paneth cell phenotype in Japanese CD subjects was similar to a North American cohort (33% vs. 26%; \( P = 0.26 \)) (Figure 2B). We also examined potential correlations between the clinical data obtained for the Japanese CD subjects and Paneth cell phenotypes (Table 1). Of note, there was no significant difference in the demographics and clinical phenotype (Montreal classification) between patients with type I and type II Paneth cell phenotypes at time of resection. In addition, we also examined whether Paneth cell phenotype and/or individual morphology category correlated with the presence or absence of granuloma. We found that while there was a trend in that the percentage of diminished Paneth cell morphology inversely correlated with the presence of granuloma as we have shown previously (19), overall there was no significant correlation between Paneth cell phenotype or each Paneth cell morphology category and granuloma (Supplemental Figure 2).

Type I Paneth cell phenotype was associated with poor clinical outcome in Japanese CD. To determine whether Paneth cell phenotype also correlated with clinical outcome in patients undergoing resections in Japanese CD, we analyzed the clinical outcome only for those patients who had received postoperative prophylactic therapy, as postoperative prophylaxis has been shown to be associated with outcome (26). We listed the characteristics of the patients that were included for outcome analysis in Supplemental Table 1. We found that in Japanese CD patients, type I Paneth cell phenotype was also associated with shorter time to disease recurrence after resection (\( P = 0.013 \); hazard ratio = 2.10, 95% CI = 1.04–4.24) (Figure 3). This association remained significant in multivariate analysis (Supplemental Table 2), thus replicating the finding that we have previously shown in North American CD patients (19).

\( ATG16L1 \) T300A was not associated with Paneth cell defect in Japanese CD. We next examined the potential associations of type I Paneth cell phenotype with 56 SNPs, selected based on known CD susceptibility associations (3, 5, 21–23, 27–29) or known association with Paneth cell function (17) (Supplemental Table 3). These SNPs include coding variants for \( ATG16L1 \) (T300A) associated with Paneth cell defects in North American CD cohorts and in genetic mouse models (18, 19, 30). Among the Japanese CD patients with available genotype data (\( n = 98 \)), we found no significant difference between the numbers of the \( ATG16L1 \) T300A risk allele and the percentage of normal Paneth cells (\( R^2 = 0.01717; P = 0.20 \)) (Figure 4A). In contrast, in North American CD with wild-type \( NOD2 \) (to eliminate potential confounding factors from \( NOD2; n = 97 \)), the numbers of \( ATG16L1 \) T300A risk allele correlated with the percentage of normal Paneth cells (\( R^2 = 0.04387; P = 0.0395 \); Figure 4B), a finding that we have shown previously with a smaller CD cohort (18). Of note, the allele frequency of \( ATG16L1 \) T300A in our Japanese CD cohort was 40%, a level comparable to what has been published (27). Therefore, in contrast to the North American CD, \( ATG16L1 \) T300A was not associated with the percentage of normal Paneth cells or type I Paneth cell phenotype in Japanese CD. \( NOD2 \), another CD susceptibility gene previously shown to be associated with Paneth cell defect in North American CD patients, is nonpolymorphic in Japanese patients (31, 32), and thus was not included on the Japonica SNP array.
Western CD susceptibility allele LRRK2 M2397T was associated with Paneth cell defect in Japanese CD. Interestingly, the only SNP within the pool of 56 selected alleles for hypothesis-driven correlation analysis that showed significant association with Paneth cell defects in Japanese CD was LRRK2 M2397T (Figure 4C). The LRRK2 M2397T SNP (rs3761863) is a missense susceptibility allele for European ancestry CD (3, 33), and Lrrk2 knockout mice have defective Paneth cells (34). In the Japanese CD cohort, we found that the numbers of the T (risk) allele LRRK2 M2397T correlated with the percentage of normal Paneth cells ($R^2 = 0.247; P = 3.62 \times 10^{-4}$; Figure 4C). The results suggest that 2 defective copies of the LRRK2 gene product may be required to illicit Paneth cell defects in Japanese CD patients.
To determine whether this correlation was universal, we next analyzed the correlation of \textit{LRRK2} M2397T and Paneth cell defects in the North American CD cohort (with wild-type \textit{NOD2} as described above). Surprisingly, there was no association between the numbers of \textit{LRRK2} M2397T risk allele and the degree of Paneth cell defect ($R^2 = 0.02054$; $P = 0.76$; Figure 4D). As \textit{ATG16L1} T300A and \textit{NOD2} SNPs are known to be associated with Paneth cell defects in North American CD cohorts, we performed further analyses with the North American CD patients with \textit{ATG16L1} T300A and/or \textit{NOD2} risk alleles ($n = 116$ with \textit{LRRK2} M2397T status available) as well as the patients without any risk alleles for \textit{ATG16L1} T300A or \textit{NOD2} ($n = 15$ with \textit{LRRK2} M2397T status available).

There was no association of \textit{LRRK2} M2397T and the percentage of normal Paneth cells when all patients (including \textit{ATG16L1} T300A and/or \textit{NOD2} risk alleles) were analyzed ($P = 0.13$; Supplemental Figure 3A) or when only the subset of patients without \textit{ATG16L1} T300A and \textit{NOD2} risk alleles were analyzed ($P = 0.82$; Supplemental Figure 3B). However, given the relatively small sample size with wild-type \textit{ATG16L1} T300A and \textit{NOD2}, further investigation with a greater sample size is needed. Therefore, our data suggest that there are dichotomous effects of the \textit{ATG16L1} and \textit{LRRK2} susceptibility alleles on Paneth cell defects within Japanese and North American CD cohorts. Interestingly, \textit{LRRK2} M2397T itself was not associated with outcome ($P = 0.5931$) (Supplemental Table 2).

Unbiased genome-wide association identified additional candidate genes associated with Paneth cell defect in Japanese CD. We subsequently performed an unbiased GWAS analysis to identify additional potential novel SNPs that could be associated with the degree of Paneth cell defects. Although no SNPs reached genome-wide significance ($<5 \times 10^{-8}$), 45 SNPs were identified as candidates ($P \leq 5 \times 10^{-6}$), as well as 8 nonsynonymous SNPs.
Figure 5. Genome-wide association analysis results for Paneth cell defect in Japanese Crohn’s disease (CD) patients. Locus zoom plots of P values around selected top associated SNPs from Paneth cell phenotype regression analysis. The top associated SNPs for (A) ZBTB16, (C) MAFB, and (E) FER are shown as purple diamonds and the remaining SNPs are shown as circles, with color indicating the level of linkage disequilibrium ($R^2$) with lead SNP. (B, D, and F) Corresponding normal Paneth cells–numbers of risk alleles correlation. Error bars represent ± SEM. (B) $P = 2.74 \times 10^{-6}$; (D) $P = 5.11 \times 10^{-7}$; (F) $P = 7.22 \times 10^{-7}$; all by linear regression.
that had $P$ values less than or equal to $5 \times 10^{-4}$ (Supplemental Figure 4). The candidate SNPs were categorized into 9 gene regions (Table 2 includes the 9 top-associated SNPs in these regions). Among the 9 candidate genes, ZBTB16, MAFB, and FER (Figure 5) were of particular interest. ZBTB16 modulates autophagy by degrading Atg14 (35). MAFB is an important transcriptional factor for macrophage differentiation (36), and FER plays a role in neutrophil chemotaxis (37), both of which are involved in innate immunity and plausibly affect autophagy functions (38, 39). Therefore, genes involved in autophagy and innate immunity are candidates for the development of Paneth cell defect in Japanese CD. Further studies in additional cohorts are needed to confirm or refute these findings. Supplemental Table 4 shows the top functional or nonsynonymous SNPs associated with Paneth cell defect ($P \leq 5 \times 10^{-4}$).

Pathway analysis revealed association with autophagy and inflammatory cytokines in candidate genes associated with Paneth cell defect. To determine the spectrum of pathways involved in the genes associated with Paneth cell defect in Japanese CD, we first performed gene enrichment, protein interaction, and KEGG pathway analyses to explore the biological functions of the top annotated genes corresponding to SNPs associated with abnormal Paneth cells with $P$ values less than or equal to $1 \times 10^{-3}$ ($n = 288$) to obtain an overview of these pathways. Our results demonstrate that the gene list was enriched for specific molecular functions and biological pathways that could be related to Paneth cell phenotype. Twelve functional annotation clusters with an enrichment score of E greater than 1.4 (corresponding to unadjusted $P$ value of 0.05) were identified for this gene set (Supplemental Table 5). We also used the 288-gene list to generate a protein-protein interaction network (Supplemental Table 6). The functional annotation of the network is shown in Supplemental Figure 5 and the KEGG pathways ($P < 0.05$) for these genes are listed in Supplemental Table 7. Many of the pathways and networks, such as cell adhesion and cytoskeletal remodeling, are known to be linked to IBD pathogenesis (4).

We next focused our analysis on the 17 identified candidate genes (LRRK2 and genes described in Table 2 and Supplemental Table 4). Ingenuity pathway analysis showed that the 17 genes identified can be connected in 2 distinct networks (Figure 6). The majority are involved in a gene network regulating PI3K and ERK signaling pathways (Figure 6A). Both of these pathways can regulate the mTOR pathway, a known upstream regulator of autophagy (40, 41). Moreover, in previous studies, mTOR signaling has been linked to stem cell health and Paneth cell differentiation (42–44). A second subset of genes were linked to the TNF-α signaling
Table 2. Candidate SNPs and genes associated with Paneth cell defect in Japanese Crohn’s disease by GWAS ($P \leq 5 \times 10^{-6}$)

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Chr</th>
<th>Position* (bp)</th>
<th>Allele1/2</th>
<th>$P$ value</th>
<th>Beta</th>
<th>Gene</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12481514</td>
<td>20</td>
<td>4225920</td>
<td>A/G</td>
<td>$2.52 \times 10^{-7}$</td>
<td>4.469</td>
<td>ADRA1D</td>
<td>intronic</td>
</tr>
<tr>
<td>rs723080</td>
<td>20</td>
<td>39357465</td>
<td>C/T</td>
<td>$5.11 \times 10^{-7}$</td>
<td>-4.754</td>
<td>MAFB</td>
<td>upstream</td>
</tr>
<tr>
<td>rs147629807</td>
<td>5</td>
<td>107889038</td>
<td>-/T</td>
<td>$7.22 \times 10^{-7}$</td>
<td>-8.139</td>
<td>FER</td>
<td>upstream</td>
</tr>
<tr>
<td>rs17318450</td>
<td>5</td>
<td>35482314</td>
<td>A/G</td>
<td>$1.62 \times 10^{-6}$</td>
<td>-5.826</td>
<td>PRLR</td>
<td>upstream</td>
</tr>
<tr>
<td>rs12494894</td>
<td>3</td>
<td>158596415</td>
<td>A/G</td>
<td>$2.25 \times 10^{-6}$</td>
<td>-6.381</td>
<td>MFSD1</td>
<td>downstream</td>
</tr>
<tr>
<td>rs72622838</td>
<td>8</td>
<td>72100325</td>
<td>A/G</td>
<td>$3.04 \times 10^{-6}$</td>
<td>-4.181</td>
<td>EYA1</td>
<td>downstream</td>
</tr>
<tr>
<td>rs11978753</td>
<td>7</td>
<td>152717110</td>
<td>G/T</td>
<td>$3.65 \times 10^{-6}$</td>
<td>-5.941</td>
<td>ACTR3B</td>
<td>downstream</td>
</tr>
<tr>
<td>rs2238823</td>
<td>22</td>
<td>45936320</td>
<td>A/G</td>
<td>$3.80 \times 10^{-6}$</td>
<td>-5.626</td>
<td>FBLN1</td>
<td>intronic</td>
</tr>
</tbody>
</table>

*Positions are based on the Genome Reference Consortium human build 37 (GRCh37). Chr, chromosome.

Discussion

We have previously shown that Paneth cell phenotypes are associated with CD susceptibility alleles, pathological hallmarks, unique transcriptomic and mucosal microbiome changes, and outcome in patients undergoing surgery in North American CD (19, 20). However, it was unclear whether similar associations would be present in CD patients from different genetic/ethnic backgrounds, particularly as the incidence of CD in other ethnic groups has continued to rise (50, 51). In the current study, we first set out to determine the prevalence, and the genetic and clinical associations, of Paneth cell defect in Japanese CD, a population who possess a distinct spectrum of susceptibility genes compared with European ancestry CD. We demonstrated that not only was Paneth cell defect prevalent in the Japanese CD, clinically it also correlated with prognosis in Japanese CD patients after surgery. Surprisingly, there were dichotomous effects of ATG16L1 and LRRK2 between the 2 cohorts. We also identified several associated SNPs in genes linked to autophagy and TNF-α signaling networks in association with Paneth cell defect.

While ATG16L1 T300A was associated with defective Paneth cells in both mouse models and North American adult CD patient cohorts (18, 52, 53), this genetic association was not observed in Japanese CD. In contrast, we found that LRRK2 M2397T, a susceptibility allele for European ancestry but not for Japanese or Korean CD (3, 23, 25, 33, 54), was associated with Paneth cell defect in Japanese, but not North American CD. There are several important considerations when interpreting these findings. First, given the different allele frequencies of ATG16L1 T300A between European and Asian CD cohorts (5), it is possible that a larger sample size is required to achieve sufficient power to detect an association between ATG16L1 T300A with Paneth cell defect in Japanese CD. Second, while LRRK2 M2397T is not a susceptibility allele for Japanese CD, it could also represent a power issue in the Japanese populations. In addition, it is possible that LRRK2 M2397T, a missense variant that could result in functional defect of LRRK2, may be associated with other aspects of CD manifestations. A recent study in leprosy (a granulomatous disease) showed that patients harboring the same LRRK2 M2397T allele suffer from excessive proinflammatory responses (55). Interestingly, overlapping susceptibility loci between CD and mycobacterial infection (particularly leprosy) is established (3). A longitudinal study would provide additional insight into the clinical relevance of LRRK2 M2397T in Japanese CD. Additional advanced sequencing technologies such as fine mapping and deep pathway (Figure 6B), consistent with the central role of TNF-α in IBD pathogenesis (4, 6, 45). TNF-α has also been implicated in the homeostasis of Paneth cell function (46–49). Therefore, the candidate genes for Paneth cell dysfunction in Japanese CD patients potentially act through modulating autophagy and TNF-α signaling. We also examined the potential interactions between the 17 identified genes and known CD susceptibility genes (6). We found interactions with several IBD susceptibility genes in both networks (6). For example, IBD susceptibility genes FYN, HCK, HNF4A, MAPK1, and RPS6KB1 were involved in the PI3K network, which included genes MAFB, ADRA1D, PRLR, ACTR3B, CEP192, CAND2, FOXM1, MFSD1, and RPL11 that were identified in this study. Likewise, IBD susceptibility genes IFNG, IL6ST, LRRK2, and CEBPB, as well as genes identified in this study LRRK2, ZBTB16, FBLN1, and EYA1, were involved in the TNF-α network. Therefore, the genes identified in this study potentially act in concert with known IBD susceptibility genes in autophagy and TNF-α signaling networks in association with Paneth cell defect.
sequencing of the \textit{LRRK2} gene and expression quantitative trait loci analysis will be important to validate the association of the M2397T SNP with Paneth cell defect, as has been shown in other studies (22, 29).

Using mouse models, we previously showed that Paneth cell defect is the result of gene and environment interactions, suggesting that proper environmental triggers or genetic context may be required to illicit Paneth cell defect in CD patients. The sharp contrast of divergent correlations of \textit{ATG16L1} T300A and \textit{LRRK2} M2397T with Paneth cell defects in North American and Japanese CD cohorts suggests that the environmental factors that the North American and Japanese CD patients harboring these variants encountered could be distinct. Furthermore, based on our current understanding of the autophagy machinery, in complex biologic processes such as Paneth cell function and intestinal homeostasis, the functions of these autophagy-associated genes may need to be orchestrated in distinct fashions depending on the environmental insults unique to these populations (56). One important direction is to use Paneth cell phenotype analysis as a platform to identify potential environmental factors or additional genetic factors that interact with specific host mutants in triggering Paneth cell defect in CD patients and mouse models with \textit{LRRK2} deficiency. In addition, environmental and genetic factors that could exert a protective effect of Paneth cell defect should be investigated through large-scale studies. Also of particular interest is to compare the Paneth cell phenotype in first- and second-generation Asian immigrants to North America (and vice versa), an approach that has yielded insight into disease pathogenesis in epidemiological studies (57–60). Based on our finding, we predict that while the genetic landscape of the second-generation immigrants will be similar to their parents, the genetic correlation to Paneth cell defects will trend more toward European ancestry North American CD. In addition, conducting genetic–Paneth cell defect analysis with focused ethnic groups (e.g., Ashkenazi Jewish) (61, 62) and with cross-ethnic groups (5) may provide additional insight into potential gene–gene interactions in triggering Paneth cell defect.

Among the candidate genes shown to be associated with Paneth cell defects by hypothesis-free association, \textit{ZBTB16}, \textit{MAFB}, and \textit{FER} have been shown to be linked to autophagy and components of the innate immune response that could be modulated with autophagy. \textit{ZBTB16} has been shown to regulate autophagy by mediating the proteasomal degradation of Atg14L (35). \textit{ZBTB16} is also involved in type 2 innate lymphoid cell function (63), NK cell differentiation (64), and regulation of inflammatory signaling (65). \textit{MAFB} is associated with macrophage differentiation (36, 66). \textit{FER} plays a role in leukocyte recruitment and intestinal barrier function in response to bacterial lipopolysaccharide recognition (67). Both macrophage differentiation and leukocyte recruitment are important elements of innate immunity and involves autophagy (38, 39). Thus, the observation that genes involved directly and indirectly in autophagy were candidates for Paneth cell defect supports the notion that autophagy is a central pathway that controls the intestinal homeostasis (68).

We and others have found that molecular changes in processes such as autophagy and TNF signaling converge to affect Paneth cell function in CD patients and mouse models of intestinal inflammation (19, 45, 52, 69, 70). This in turn correlates with the development of dysbiosis and inflammation (20). Together, these studies suggest that, at least in a subset of CD patients, Paneth cells are the origin of intestinal inflammation (10). Future studies that examine the transcriptomic and signaling pathways of Paneth cells from patients with different genetic backgrounds will provide additional insight.

Importantly, although the SNPs for Paneth cell defect identified in this study were largely distinct from the known SNPs for European ancestry CD, Paneth cell phenotype still correlated with prognosis in Japanese CD patients undergoing surgery. These results indicate that Paneth cell phenotype, as an integrated readout for combinatorial effect of host genetics and environmental factors, could potentially be applied as a universal prognostic biomarker for CD patients of different ethnicity/genetic backgrounds undergoing resection. Of note, a recent study has highlighted that the genes associated with CD susceptibility and prognosis are distinct (71). Therefore, development of a potential gene score that predicts postoperative recurrence may be complementary to using Paneth cell phenotype as biomarker. In addition, while we did not observe the correlation between Paneth cell phenotype and the presence of granuloma, we did observe a trend between diminished Paneth cells and granuloma, a finding that was significant in our previous study in North American CD (19). Therefore, future studies with a larger sample size may provide more insight.

In summary, we demonstrated that Paneth cell defect was prevalent in Japanese CD, was associated with different spectrum of genes compared with North American CD, and was predictive of prognosis. The genes involved in Paneth cell defect in Japanese CD mainly affect autophagy and TNF signaling pathways. Paneth cell phenotype can be applied as a universal clinical and biological relevant biomarker for CD patients with diverse genetic backgrounds.
Methods

Subjects. A total of 110 Japanese patients with CD were included in this study. All patients were diagnosed with CD and underwent ileal resection at Tohoku University Hospital (Sendai, Japan) between 2003 and 2014. The self-reported ethnicity of all patients was Japanese. The diagnosis of CD was made based on clinical symptoms and endoscopic, radiographic, and histological findings according to conventional criteria proposed by the Japanese Ministry of Health, Labour and Welfare (72). Demographics and clinical parameters (including clinical phenotype by Montreal classification, medication received prior to and after surgery, time to disease recurrence after resection) were extracted from the medical records blinded to Paneth cell phenotypes.

In addition, a total of 164 CD patients who underwent ileal resection were recruited at Washington University School of Medicine between 2005 and 2013, including a subset of patients (n = 50) that were previously reported (19).

Paneth cell phenotype analysis. Ileal resection samples previously collected as part of standard diagnostic procedures and stored in Pathology Department archives at Tohoku University and Washington University were used for Paneth cell phenotype determination. The specimens were processed as formalin-fixed, paraffin-embedded (FFPE) tissue blocks stored at 24°C as per routine surgical pathology practice. To be included in the study, the proximal margin tissue resection samples must have contained at least 50 well-oriented intestinal crypts (73) as determined by a pathologist (T.C.L.). Paneth cell phenotype was determined by defensin-5 immunofluorescence, using defensin-5 antibody (clone 8C8; dilution 1:2,000; Novus Biologicals, catalog NB1110) followed by donkey anti-mouse Alexa 488 antibody (dilution 1: 500; Thermo Fisher Scientific, catalog A-21202) as previously described (20). For a subset of cases, containing with lysozyme immunofluorescence was performed using lysozyme antibody (clone C-19; dilution 1:100; Santa Cruz Biotechnology, catalog sc-27958) followed by donkey anti-goat Alexa 594 antibody (dilution 1: 500; Thermo Fisher Scientific, catalog A-11058). Each Paneth cell was classified as normal or abnormal based on the morphology of the cytoplasmic granules (19, 20, 73). For prognostic correlation, the overall Paneth cell phenotype for each patient was then defined as following: type I Paneth cell phenotype was defined as greater than or equal to 20% abnormal Paneth cells, whereas type II Paneth cell phenotype was defined as less than 20% abnormal Paneth cells (19, 20, 73). The cutoff of 20% was based on the analysis of 106 non-IBD cases, which showed that the mean percentage of normal Paneth cells was 80.63 (± 1.17 SEM). The cutoff was used in our previous studies involving North American CD cohorts (19, 20, 73). Paneth cell phenotype analysis was performed by a pathologist (T.C.L.) who was blinded to the identification and clinical phenotype of the cases.

Definitions of clinical remission and clinical relapse. Time from index ileal resection surgery of CD to relapse was calculated. Endoscopy and/or computed tomography were performed when the patients had symptoms indicating recurrence (i.e., abdominal pain, worsening of diarrhea, fever, body weight loss, nausea, and appetite loss), or at least once per year if the patients had no symptoms.

Definitions of recurrence were made by endoscopy and/or radiology. Endoscopic recurrence was defined as postoperative Rutgeerts score of i2 or more. Radiologic recurrence was defined by computed tomography showing bowel wall thickening (defined as a thickness > 5 mm) or enhancement, fistulas, intra-abdominal abscess, and bowel obstruction. All computed tomography scans were reviewed by 2 radiologists. Patients who received immunomodulators (thiopurines), biologics (infliximab or adalimumab), or elemental diet (more than 600 kcal/day) (74) after resection were considered to have received postoperative prophylaxis.

Genotyping. For Japanese CD patients, genomic DNA was obtained from peripheral blood leukocytes by standard phenol–chloroform extraction and precipitation or by utilizing an NA1000 Automated Nucleic Acid Extraction Machine (Kurabo) or PAXgene DNA Kit (BD). The genome-wide SNP genotypes of the Japanese CD patients were determined by the Japonica array, a SNP array designed specifically for the Japanese population (75). The array contains 659,253 SNPs, including tag SNPs for imputation, as well as SNPs related to phenotypes from previously reported GWAS and pharmacogenomics studies. As a part of quality control (QC) measures, SNPs with call rate less than 95% and samples with genotyping rates less than 95% were excluded from further analysis. Following SNP and sample QC, genotype data of 643,496 SNPs from 98 CD cases were available for further analysis. For North American CD patients, genomic DNA was extracted from whole blood (n = 50) as described above or from FFPE tissue using a QIAamp DNA FFPE Tissue kit (QIAGEN) (n = 114). For cases where whole blood was used for genomic DNA extraction, genotyping was performed with Immunochip as previously
described (19). For cases where FFPE tissue was used for genomic DNA extraction, genotyping was performed with Taqman SNP genotyping assays (Thermo Fisher Scientific).

**Imputation.** Untyped genotypes were imputed in the GWAS samples using IMPUTE2 (version 2.3.2) and 1,070 healthy individuals from Japan (1KJPN panel, which contains > 20 million SNPs) (76) as a reference dataset. SNPs with low imputation quality (with a posterior probability score of < 0.90), minor allele frequency less than 0.10 and Hardy-Weinberg equilibrium P value less than 1 × 10⁻³ were excluded. After exclusion, genotype data of 4,198,245 SNPs from 98 Japanese CD cases were used for further analysis.

**Pathway and network analysis.** Gene enrichment and network analysis was performed on the top 288 annotated genes corresponding to the SNPs with P less than or equal to 1 × 10⁻³ (from linear regression) using online bioinformatics tools DAVID (https://david.ncifcrf.gov/tools.jsp) and STRING (http://www.string-db.org/). Functional annotation clusters corresponding to the gene list were identified using DAVID. Protein-protein interaction (PPI) network for the 288 genes was extracted using STRING. An annotation network was constructed by combining function-based information with PPI network information and visualized using Cytoscape (www.cytoscape.org). Corresponding pathways (P < 0.05) were annotated using the enrichment analysis tool in STRING. Core pathway analysis of the 17 candidate genes was performed using Ingenuity Pathway Analysis software (QIAGEN).

**Statistics.** The percentages of normal Paneth cells between North American CD and Japanese CD were compared by unpaired t test, and the prevalence of type I Paneth cell phenotype between the 2 groups was compared using Fisher's exact test. For genotype–Paneth cell defect correlation, we first adopted a hypothesis-driven approach with selected SNPs (Figure 1). These SNPs include: (a) 45 SNPs shown to have CD-specific susceptibility from an ImmunoChip study examining European ancestry CD (3); (b) an additional 38 SNPs reported in European and East Asian ancestry CD (5); (c) 14 SNPs associated with East Asian (Japanese and/or Korean) CD through GWAS analyses (21, 23, 28, 77) and deep-resequencing analysis (22); and (d) 2 functional CD-associated variants that were either reported to be associated with Paneth cell defect in our previous studies (18, 19), or whose genes involved had been implicated to be associated with Paneth cell biology in preclinical studies (17, 29). By excluding duplicates, a total of 73 SNPs were compiled as candidates. Among the 73 SNPs, 14 SNPs were very rare (minor allele frequency < 1.0%) or nonpolymorphic (i.e., NOD2 variants) in East Asian, and 3 SNPs were not constructed/included on the Japonica array. After excluding these 17 SNPs, a final 56 SNPs were selected for the hypothesis-driven approach. After Bonferroni correction, SNPs with P values less than 8.93 × 10⁻⁴ were considered significant. For hypothesis-driven analyses, we correlated the genotypes with the percentage of normal Paneth cells by linear regression.

For the subsequent unbiased, GWAS analysis (Figure 1), linear regression was performed using PLINK v1.07 software (78). SNPs with P values less than 5 × 10⁻⁸ were considered genome-wide significant. SNPs with P values less than 5 × 10⁻⁶ and nonsynonymous SNPs with P values less than 5 × 10⁻⁴ were considered candidates with nominal significance.

SNPs located within 100 kbp were considered to be in 1 region. Genes within 200 kbp of candidate SNPs were investigated. Correlation with time to disease recurrence after surgery was performed using the log-rank test and Cox-proportional hazards model. Manhattan plots were generated using the R package qqman, regional association plots were generated using LocusZoom application (79). All statistical analyses, except genome-wide linear regression, were performed using JMP 11 (SAS Institute Inc.), GraphPad Prism (version 6.5), or R software (version 3.1.3).

**Power calculation.** Using the software Quanto, we calculated the power of the current study to detect genetic variants associated with the Paneth cell phenotype, and demonstrated that in the Japanese cohort, we have a power of 80.66% to detect a variant with variance contribution of 0.07 at 0.05 significance threshold for the hypothesis-driven approach. In the Washington University cohort, we have a power of 82.65% to detect a variant with variance contribution of 0.05 with a significance threshold of 0.05. At genome-wide significance threshold (5 × 10⁻⁸), the Japanese cohort has a power of 56.91% to detect a variant with variance contribution of 0.247.

**Study approval.** The study was approved by the Ethics Committee of Tohoku University Graduate School of Medicine under protocol number 2013-1-539 and the IRB of Washington University School of Medicine under protocol number 201209047. Subjects provided informed consent prior to their participation in the study.

**Data depository.** The accession number for the genotyping data deposition at NCBI Gene Expression Omnibus repository is GSE90102.
Author contributions
TCL, TN, DPBM, Y. Kakuta, and TSS designed the study. TCL, TN, ZL, TH, Y. Kawai, MN, and Y. Kakuta acquired data. TN, TS, and TK recruited patients. TCL, TN, ZL, KLV, TH, DL, DPPM, Y. Kakuta, and TSS analyzed data. TCL, TN, DPBM, Y. Kakuta, and TSS drafted the manuscript. TN, KE, and Y. Kinouchi provided samples.

Acknowledgments
The study was supported by the Helmsley Charitable Trust (D.P.B.M., T.S.S.), Doris Duke Charitable Foundation grant 2014103 (T.C.L.), Crohn’s and Colitis Foundation grant 274415, NIH grants 1R56DK095820 (D.P.B.M., T.S.S.) and K01DK109081 (K.L.V.), NIH/National Center for Advancing Translational Sciences (NCATS) grant U1 TR000448 (T.C.L.), and Japan Society for the Promotion of Science KAKENHI grants JP15H04805 and JP15K15284 (Y.K.). Computational resources for genotype imputation were provided by the Tohoku Medical Megabank Organization supercomputer system.

Address correspondence to: Thaddeus S. Stappenbeck or Ta-Chieng Liu, Department of Pathology and Immunology, Washington University School of Medicine, 660 South Euclid Avenue, CB 8118, St. Louis, Missouri 63110, USA. Phone: 314.362.4214; E-mail: stappenb@wustl.edu (T.S. Stappenbeck). Phone: 314.747.0343; E-mail: tliu27@wustl.edu (T.C. Liu). Or to: Yoichi Kakuta, Division of Gastroenterology, Tohoku University Graduate School of Medicine, 1-1 Seiryo, Aoba, Sendai, 980-8574, Japan. Phone: 81.22.717.7171; E-mail: ykakuta@med.tohoku.ac.jp.


