HHS Public Access

Author manuscript

Gastroenterology. Author manuscript; available in PMC 2017 October 01.

Published in final edited form as:

Gastroenterology. 2016 October ; 151(4): 710–723.e2. doi:10.1053/j.gastro.2016.06.045.

A Frameshift in CSF2RB Predominant Among Ashkenazi Jews Increases Risk for Crohn's Disease and Reduces Monocyte Signaling via GMCSF

A full list of authors and affiliations appears at the end of the article.

Abstract

Background & Aims—Crohn's disease (CD) has the highest prevalence in Ashkenazi Jewish populations. We sought to identify rare, CD-associated frameshift variants of high functional and statistical effects.

Methods—We performed exome-sequencing and array-based genotype analyses of 1477 Ashkenazi Jewish individuals with CD and 2614 Ashkenazi Jewish individuals without CD (controls). To validate our findings, we performed genotype analyses of an additional 1515 CD cases and 7052 controls for frameshift mutations in the colony stimulating factor 2 receptor beta common subunit gene (*CSF2RB*). Intestinal tissues and blood samples were collected from patients with CD; lamina propria leukocytes were isolated and expression of *CSF2RB* and GMCSF-responsive cells were defined by mass cytometry (CyTOF analysis). Variants of *CSF2RB* were transfected into HEK293 cells and expression and functions of gene products were compared.

Results—In the discovery cohort, we associated CD with a frameshift mutation in *CSF2RB* (P=8.52×10–4); the finding was validated in the replication cohort (combined P=3.42×10–6). Incubation of intestinal lamina propria leukocytes with GMCSF resulted in high levels of phosphorylation of STAT5 and lesser increases in phosphorylation of ERK and AKT. Cells co-transfected with full-length and mutant forms of CSF2RB had reduced pSTAT5 following

*To whom correspondence should be addressed. **The corresponding author's contact information:** Judy Cho, Hess CSM Building Floor 8th Room 118, 1470 Madison Avenue, New York, NY 10029, TEL. (212) 824-8940, FAX. (646) 537-9452, Judy.cho@mssm.edu.

†indicates equal contributions

The authors whose names are listed below declare that they have no affiliations with or involvement in any organization with financial or non-financial interest in the subject matter or materials discussed in this manuscript.

Supplementary Materials

Supplemental figure 1: Discovery phase principal components analysis
Supplemental table 1: Summary of discovery and replication cohorts
Supplemental table 2: Power calculations
Supplemental table 3: CyTOF labeling panel

Author contributions: C.L-S, V.N., P.I., and C.JH. design this study, and contributed in writing, interpretation of data and quality assessment. H.KY. performed all the genetic statistical analysis and contributed in study design, writing, and quality assessment. B.SR. and M.DP. contributed in study design, interpretation of data, and quality assessment. D.MJ. contributed in study design, statistical analysis, and interpretation of data. R.JI., D.RJ., A.MT., A.G., P.I., K.S., H.H., M.JL., L.T., D.A., P.V. and S.MS. contributed in study design and quality assessment. M.A., R.A., H.NY., S.AW., and M.M. contributed in study design and interpretation of data. H.T., L.S., and Z.H. statistical analysis and quality assessment. L.AP. contributed in statistical analysis and interpretation of data. A.C. contributed in writing and quality assessment. R.R. and C.S. contributed in study design. L.D., R.M., S.ER., G.A., B.BM., and N.K. contributed in statistical analysis. M.AM. contributed in interpretation of data. N.SE., Z.W., F.J-A., L.T., F-H. H., S.LP., S.Y., L.P., G.J., K.A., S.B., S.JM., D.LW., K.S., C.AS., B.N., P.N., D.MJ., T.S., T.K., and S.EJ. contributed in data quality assessment.

stimulation with GMCSF, compared to cells transfected with control CSF2RB, indicating a dominant negative effect of the mutant gene. Monocytes from patients with CD who were heterozygous for the frameshift mutation (6% of CD cases analyzed) had reduced responses to GMCSF and markedly decreased activity of aldehyde dehydrogenase; activity of this enzyme has been associated with immune tolerance.

Conclusions—In a genetic analysis of Ashkenazi Jewish individuals, we associated CD with a frameshift mutation in *CSF2RB*. Intestinal monocytes from carriers of this mutation had reduced responses to GMCSF, providing an additional mechanism for alterations to the innate immune response in individuals with CD.

One Sentence Summary

Crohn's disease associated frameshift mutation in *CSF2RB* results in decreased STAT5 activation in a dominant negative manner.

Keywords

IBD; inflammatory bowel disease; risk factor; ethnic variation

Background and Aims

Genome-wide association studies (GWAS) have identified over 200 loci associated with inflammatory bowel disease (IBD), largely represented by common variants of modest effect. Taken together, these loci account for 13.1% and 8.2% of the phenotypic variance for Crohn's disease (CD) and ulcerative colitis (UC), respectively[1]. For only 10–15% of these loci, the maximally associated marker is predicted to cause a protein-coding change[2]; most association signals are believed to be driven by genetic polymorphisms which modulate gene expression, often in subtle and context-specific ways.

The large majority of IBD loci confers risk to both CD and UC and implicates numerous pro- and anti-inflammatory pathways (e.g. interleukin 23, prostaglandins, interleukin 10) and immune cell subsets[1, 2]. As with many chronic immune-mediated diseases, in ulcerative colitis, the most significant association is to the major histocompatibility complex (MHC), specifically class II alleles. In marked contrast, MHC associations are not dominant in European ancestry CD; rather, protein-coding polymorphisms in the *NOD2* (nucleotide oligomerization domain 2) gene and autophagy pathways (e.g. *ATG16L1*) involved in innate immunity represent major association signals. The *NOD2* associations confer the largest statistical effects (i.e. odds ratios) observed in IBD[2]; most prominent among these is a cytosine insertion, resulting in a frameshift mutation[3, 4] which results in impaired innate immunity, specifically a complete loss of response to microbial stimuli (e.g. bacterial peptidoglycan)[5]. CD-associated *NOD2* risk alleles are relatively uncommon, with allele frequencies in controls of less than 5%; therefore, only a minority of CD patients carry *NOD2* risk alleles. Taken together, these findings established that a significant fraction of CD patients have impaired innate immune function, that is, impaired response to microbial stimulation with *NOD2* risk allele carriage.

The extent to which 'Mendelian-like', rare protein-coding mutations with greater statistical and functional effects contribute to complex human disease is currently unknown. Compared to common variants, there is decreased power to detect association with rare variants[6]. Furthermore, rare insertion-deletion frameshift mutations are under-represented in commercially available genotyping chips due to suboptimal design features. Finally, the elucidation of rare disease associated variants may be hindered by their greater population specificity. Of particular interest here, the Ashkenazi Jewish (AJ) population has a 4–7 fold higher prevalence of CD[7, 8]. As a founder population, genetic analyses of Jewish populations have identified numerous mutations unique to the Ashkenazim, contributing to a wide-array of Mendelian diseases[9]. Importantly, while such uncommon risk alleles may be more identifiable in AJ populations, they can highlight pathways of broad significance. For example, the I1307K polymorphism in the *APC* (adenomatous polyposis coli) gene has a carrier rate of 6.1% in AJ populations and confers an odds ratio of 1.9 fold for colorectal cancer (CRC)[10]. While this specific variant is uniquely observed in AJ populations, the APC pathway is of broad significance in CRC pathogenesis, both in the form of extremely rare, highly penetrant APC mutations associated with the familial adenomatous polyposis syndrome, as well as somatic mutations in sporadic CRC[11].

In this study, we sought to identify AJ specific or predominant rare mutations of high functional effects, with a particular focus on insertion-deletion variants, which result in frameshift changes. We performed exome sequencing of AJ CD cases, and genotyped 224 frameshift variants using custom array genotyping in an AJ case-control cohort. Beyond the *NOD2* frameshift, the most significantly associated frameshift mutation was observed in *CSF2RB*, the beta-chain receptor for the GM-CSF, IL-3 and IL-5 cytokines. This association was strongly replicated in an independent case-control cohort. We observe a marked decrease in GM-CSF signaling in monocytes from heterozygous frameshift carriers, further implicating innate immune deficiencies in the pathogenesis of CD. Our study highlights the value of the AJ population in identifying disease-associated coding region polymorphisms of high statistical and functional effects in complex diseases and further implicates impaired innate immunity in CD pathogenesis.

Methods

Human subjects

Written informed consent for genetic analyses was obtained from all subjects through the respective Institutional Review Boards from the contributing institutions (Supplemental table 1). Resected terminal ileal tissues and blood samples from Crohn's patients were obtained after written informed consent under the guidelines of the Human Investigations Committee (IRB) of the Icahn School of Medicine at Mount Sinai. IBD patients had diagnoses confirmed at each recruiting site by a health care provider, based on standard criteria including clinical presentation, as well as endoscopic, radiologic and/or pathologic confirmation.

Genetic analysis

We conducted exome sequencing on 47 AJ CD patients and 3 controls with full AJ ancestry as confirmed by GWAS principal component analysis (PCA). Exome capture was performed using the Agilent SureSelect library and sequencing performed using paired-end 75-bp reads with the Illumina GAII. Sequence reads were mapped to the reference genome (hg18) using the BWA program using default parameters[12], and coordinates were translated to the hg19 reference genome using LiftOver in the UCSC Genome Browser with default parameters[13]. The Genome Analysis Toolkit (GATK) v2 was used to call alleles at variant sites[14]. For the discovery case-control association study, a total of 7,813 individuals with Crohn's disease and 8,781 unaffected controls, of which 1,908 and 2,955 have Jewish ancestry, respectively, were enrolled in this study at centers throughout North America, Europe, and Israel.

Markers identified by exome sequencing not present on the Illumina HumanExome Beadchip and passing design feature filters were added. Genotyping data were generated at three genotyping centers (Philadelphia, PA, Manhasset, NY, and Los Angeles, CA) using the same custom genotyping array. The data were combined and preliminary genotypes were called jointly using input from all three centers in GenomeStudio. Following guidelines produced by the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) consortium, we enforced quality control using SNP metrics based on fluorescent probe intensities and genotype frequencies, as well as visual inspection of markers with uncertain genotyping quality[15].

Following manual review, additional quality thresholds for sample exclusion (genotype call rate < 0.968, $p50_{GC}$ < 0.758, or average heterozygosity > 0.31) were applied. Related samples were identified using pairwise identity-by-descent detection in PLINK[16] and a minimum necessary set of individuals was removed to yield a cohort without first-or second-degree genetic relationships. Samples with discrepancy between self-reported gender and genotypic gender were excluded. Following association analysis, cluster plots of the frameshift mutations were re-examined to ensure high-quality genotype calling.

We created a set of 10,312 null independent autosomal polymorphisms by removing markers with pairwise linkage disequilibrium (LD) of $r^2 > 0.05$, those with minor allele frequency (MAF) below 0.05, custom content, and variants within established IBD loci. PCA was performed using the principal component function in R version 2.15.1 (Supplemental figure 1). The PCA was conducted on our dataset in conjunction with a reference cohort of non-Jewish European-ancestry samples on which exome chip data was available. Using membership in the PCA cluster as genetic validation of AJ ancestry (Supplemental figure 1, red dots), we conducted all further analyses using only samples with 100% AJ ancestry: 1,477 CD cases and 2,614 independent healthy controls (Table 1).

For the exome sequencing experiment, we performed power calculations using binomial distribution probability estimates, in order to assess the proportion of variation in the genome captured in our study (Supplemental table 2). We also calculated power estimates for the Discovery Phase association testing using cumulative probability densities under a normal distribution.

To corroborate discovery findings, single-point genetic association was done with an independent AJ replication cohort containing 1,515 CD cases and 7,052 healthy controls. Ashkenazi Jewish ancestry was validated for all samples by principal components analysis using genome-wide variation. Association testing was conducted using χ^2 tests; pooled analyses were performed on the summed allele counts of both the Discovery and Replication Phase genotypes (Table 1).

Single cell suspensions from ileal resection tissues

Resection specimens were washed in HBSS (Life Technologies 14170112) and then incubated in HBSS (Ca^{2+} and Mg^{2+} free) with 1mM 2-mercaptoethanol (Millipore es-007e) for 10 min at room temperature. Samples were transferred to HBSS (Ca^{2+} and Mg^{2+} free) with 5mM EDTA and 10mM HEPES buffer (Life Technologies 15630080) for 20 min in 37 °C shaker, minced and placed in HBSS (Ca^{2+} and Mg^{2+}) with 2% heat-inactivated (HI) FBS, Collagenase (Sigma C5138) and DNase I (Sigma DN25) for 30 min into a shaker set at 100 rpm and 37 °C . The sol ution was vortexed and filtered through a 70-micron filter. Single cell suspension was centrifuged and lamina propria leukocytes were collected by a 40%:80% Ficoll gradient for 20min at 2500 rpm and 20°C. Lamina propria leukocytes were collected at t he interphase and washed with FACS buffer (PBS containing 2% HI-FBS and 5mM EDTA). Cells were labeled with 5μM cisplatin for 2 minutes to identify dead cells (Fienberg HG, Cytometry Part A, 2012) and live cells counts were determined using a Neubauer counting chamber and a light microscope. Half of the cells were treated with 10 ng/ml GM-CSF for 15 minutes, while the other half were left unstimulated. All cells were then fixed with 1.6% formaldehyde for 20 minutes on ice and processed for Flow Cytometry and CyTOF analysis.

CyTOF analysis

All mass cytometry reagents were purchased from Fluidigm Inc. (formerly DVS Sciences) unless otherwise noted. Stimulated and non-stimulated cell suspensions were first barcoded using unique combinations of CD45 antibodies conjugated to distinct isotopes. The barcoded samples were then pooled and stained with a panel of antibodies against cell surface markers (Supplemental table 3) for 30 minutes on ice, then washed, permeabilized with ice-cold 100% methanol, and stained with antibodies against intracellular phospho-protein epitopes (Supplemental table 3). After antibody staining, the cells were washed and incubated in 1.6% formaldehyde with 125nM nucleic acid Ir-intercalator to label all nucleated cells. Immediately prior to acquisition, the cells were washed in PBS, then in distilled water and resuspended in distilled water containing a 1/20 dilution of EQ 4 Element Normalization beads. The samples were acquired on a CyTOF2 following routine instrument tuning and optimization. The resulting FCS files were normalized and concatenated using CyTOF acquisition software and uploaded to Cytobank[17] for analysis. FCS files were manually pre-gated to exclude EQ beads, cell debris, cell doublets and dead cells, and the resulting live CD45+ population was clustered (Table S3, clustering markers designated in blue) using SPADE[18]. Putative populations were identified on the basis of multidimensional marker coexpression. Population frequencies were calculated as a percentage of total CD45+ events, and population distributions of CSF2RA (CD116) and CSF2RB (CD131) were determined for each population, normalized to non-stained controls for the corresponding

mass channels. Phospho-STAT5, -AKT and -ERK expression was normalized to non-stained controls for the corresponding mass channel and expressed as a fold change between GM-CSF stimulated samples and non-stimulated controls. The Cytof data for phosphorylated STAT5 are normally distributed, thus the *P* values were calculated with two sample t tests having equal variance. Other *P* values were generated with the non-parametric Wilcoxon rank sum test.

Clone construction and transfection studies

The complete coding regions of CSF2RA, CSF2RB_WT and CSF2RB_Mut were synthesized by GeneArt gene synthesis and cloned into the Gateway pDONR™221 Vector (GeneArt, Life Technologies), and the clone sequences confirmed by Sanger sequencing. For immunoblot analyses, CSF2RB_WT and Mut were sub-cloned into the Gateway destination vector pcDNA6.2_C-EmGFP-DEST (Life-Technologies), and CSF2RA into pcDNA6.2_C-YFP-DEST (Life-Technologies). For co-localization studies, CSF2RA and CSF2RB_WT were sub-cloned into PCDNA6.2/V5-DEST (Life-technologies) and pDEST-mcherry-N1 (Addgene plasmid # 31907), respectively.

HEK293 cells (ATCC CRL 1573) were seeded at 2×10^5 cells per well in 24 well plates and transfected with 500ng of total DNA with Lipofectamine 3000 (Life Technologies). Transfected DNA was comprised of combinations of pcDNA6.2_C-YFP-DEST -CSF2RA, pcDNA6.2_C-EmGFP-DEST-CSF2RB_WT, and pcDNA6.2_C-EmGFP-DEST-CSF2RB_Mut (Life Technologies). Twenty-four hours post-transfection, cells were stimulated with 0, 0.4, 2 or 10 ng/ml of GM-CSF for 15 min and lysed in RIPA buffer (Millipore). For co-localization studies, CSF2RA, CSF2RB_WT and CSF2RB_Mut were co-transfected and stimulated with GM-CSF (10 ng/ml).

Immunoblotting and confocal analysis

Total protein samples of transfected HEK293 or isolated human monocytes were prepared for SDS-PAGE. From each sample, 10–25 μg of protein was separated on a Mini-Protean TGX 4–15% (Bio-Rad) gel and transferred to a PVDF membrane using a Bio-Rad electrophoretic transfer apparatus (Trans-Blot Turbo, Bio-Rad). After blocking with 5% BSA, the membrane was probed with either of the following primary antibodies: a 1:1000 GFP (Abcam ab1218), 1:10000 dilution of Actin (Sigma A5441), 1:500 of STAT5 (Santa Cruz sc-835), 1:1000 of pSTAT5 (Cell Signaling 4322), 1:1000 of pERK1/2 (Cell Signaling 9101) and 1:1000 of pAKT (Cell Signaling 4058). A secondary incubation of either 1: 40000 dilution of horseradish peroxidase (HRP) conjugated goat anti-mouse or 1:20000 dilution goat anti-rabbit secondary antibody was applied to the membrane. The immunoblots were developed with SuperSignal West Pico (Life Technologies) and band intensities were quantified through ImageJ and the fold-change normalized to no GM-CSF treatment.

HEK293 cells were plated on glass coverslips (Electron Microscopy Sciences) and co-transfected with CSF2RA, CSF2RB_WT and CSF2RB_Mut plasmids as previously described. Cells were fixed with 2% paraformaldehyde PBS solution (Sigma), and incubated with rabbit anti-CSF2RB C-terminal antibody (Santa cruz: sc-376582) at 1:200 dilution in blocking solution (0.6% saponin, 5% fetal calf serum, PBS pH7.4) at room temperature for 1

hour, and followed by Alexa 555 conjugated goat anti-rabbit secondary antibody (Life technologies) at 1:500 dilution in blocking solution at room temperature for 1 hour. All imaging was performed on a Zeiss LSM510META confocal laser scanning confocal microscope (Carl Zeiss, USA) equipped with 488nm and 565 nm laser line and 63X oil immersion objectives with 1.4 numerical aperture (NA).

ALDEFLUOR activity measurement and flow cytometry

Isolated monocytes were plated at 2×10^6 in 2ml complete RPMI (Gibco) in a 24 well plate (Corning) and cultured for 7 days at 37°C and 5% CO2. Fifty percent of the cell culture media was replaced with fresh media containing GM-CSF (10ng/ml, Peprotech) and MCSF (50ng/ml, Peprotech). On day 7, cells were harvested and washed with ice cold PBS. Cultured monocytes were stained with the ALDEFLUOR KIT (Stem cell technologies) according to the manufacturer's protocol. After washing, cells were stained with CD11c (BD Horizon, B-ly6) and CD14 (Invitrogen) and analyzed on an LSR Fortessa II (BD). FCS files were analyzed using FlowJo software.

For intracellular staining, surface stained cells were fixed for 20 minutes on ice using the BD Cytofix/Cytoperm kit (BD Biosciences). Intracellular staining with human CD131-Pe antibody was performed using Perm/wash Solution (BD Biosciences). Cells were analyzed on a BD Fortessa II (BD Biosciences).

Monocytes isolation from *CSF2RB* frameshift mutation carriers

Experiments were performed with monocytes from three pairs of CD patients, each time including a CSF2RB frameshift carrier and a non-carrier. The carrier and control pair were matched based on gender and age, within a +/− ten year gap. The five tubes of blood were drawn in 10mL sodium heparin tubes within one hour between carrier and control patients. PBMCs were obtained from blood through a density gradient centrifugation with Ficoll-Paque Plus (GE) and Leucosep tubes. Monocytes were isolated from PBMCs with the Monocyte Isolation Kit II (Miltenyi) according to company protocol. Negatively selected cells were plated at 1×10^6 cells per well in 24 well plates. For short-term studies, monocytes were kept in RPMI 1640 (Life Technologies) overnight, stimulated with GM-CSF for 15 min, and lysed in RIPA buffer. Monocytes for 7-day studies were cultured in RPMI 1640, 10% fetal bovine serum, 10 ng/ml MCSF, and either 0, 0.4, 2 or 10 ng/ml of GM-CSF. Media were collected and replaced at 24, 72, and 120 hours.

Results

Exome sequencing and association studies

Exome sequencing detected a total of 371 frameshift mutations in 47 CD patients and 3 controls having full AJ ancestry[19]. Of these, 224 mutations passed the Illumina design filters for custom content and were added to the HumanExome beadchip. Using this array, we performed association analyses in 1,477 unrelated CD cases, and 2,614 independent healthy controls passing quality filters, with full AJ ancestry genetically validated using PCA (discovery cohort). Sequencing 94 independent chromosomes would have 0.39, 0.62 and 0.77 statistical power to identify variants of 0.005, 0.01 and 0.015 minor allele frequencies,

respectively (Supplemental table 2). Given the size of our discovery association cohort (Table 1), we would have reasonable power under a range of effect sizes (i.e., odds ratios) to detect nominal association (*P* value < 0.001) for variants having minor allele frequencies of 0.01 and 0.015 (Supplemental table 2). Increasingly rarer variants have progressively decreasing power to detect association; below 0.005, only limited power to detect association is predicted (Supplemental table 2). Therefore, we restricted subsequent analyses to those variants having an allele frequency > 0.005. Of the 224 frameshift mutations, 88 demonstrated a minor allele frequency of > 0.005.

Quantile-quantile (Q-Q) analysis of the 88 frameshift mutations was evaluated in order to assess the validity of the null distribution assumption. It revealed only two markers demonstrating significantly greater association evidence than predicted, namely the *NOD2* frameshift mutation and a frameshift mutation in the *CSF2RB* gene (*P* value < 8.52×10^{-4}) (Figure 1A), which encodes for the colony stimulating factor 2 receptor, beta chain. The *CSF2RB* mutation involves deletion of a single guanine base, which is predicted to cause a frameshift at codon 708 of *CSF2RB*. This would result in the expression of a 729-amino acid sequence, compared to 897 residues encoded by the canonical transcript (Figure 1B). *CSF2RB* encodes for the common beta-chain receptor for the IL-3, IL-5 and GM-CSF (granulocyte-monocyte colony stimulating factor) cytokines.

The *CSF2RB* association in the discovery cohort is nominally significant, with Bonferroni-corrected significance of 88 markers estimated at 5.68×10^{-4}. Therefore, we performed replication genotyping of *CSF2RB* in an independent case-control cohort of 1515 AJ CD cases and 7052 AJ controls (Table 1) and observed almost identical allele frequencies and odds ratios compared to the discovery cohort, with a combined *P* value of 3.42×10^{-6}. While this variant is more than an order of magnitude less frequent in non-Jewish European ancestry compared to AJ cohorts, we observed the same direction of effect (i.e., frameshift mutation frequency greater in cases, 0.0022, than in controls, 0.0016) and similar odds ratios (1.68, *P* value = 0.09) in the non-Jewish cohort.

CSF2RB and NOD2 are both expressed in intestinal innate immune cells, but demonstrate distinct patterns of expression in CyTOF-defined cell subsets

We next sought to compare the intestinal expression of *CSF2RB* and *NOD2* by CyTOF, which allows for a highly refined cellular definition from complex tissues; using antibodies labeled with heavy metals having minimal spectral overlap between neighboring isotopes, simultaneous analysis in a single panel of many proteins is possible. We designed a CyTOF panel that included immune cell classification markers, with a particular focus on innate immunity, as well as members of the GM-CSF signaling pathway (Supplemental table 3).

While both *CSF2RB* and *NOD2* function in innate immunity, key differences in the distribution of their expression are observed. Among the most prevalent cell subsets, while CSF2RB (Figure 2A) and NOD2 (Figure 2C) are both expressed in monocytes and macrophages, CSF2RB is highly expressed in eosinophils, where NOD2 expression is nominal (n=3, *P* value =0.0002). Conversely, high levels of NOD2, but not CSF2RB expression are observed in plasma cells (n=3, *P* value = 0.004). NOD2 is broadly expressed in monocytes, macrophages and dendritic cells (Figure 2C). In contrast to NOD2, more

restricted expression of CSF2RB within mononuclear phagocytes is observed; higher expression of CSF2RB was observed specifically in HLA-DR high expressing monocytes from inflamed tissues (Figure 2A). The HLA-DR high, CSF2RB positive monocyte population in inflamed tissue also expresses CD206, a marker for resident macrophages[20] (Figure 2D). No expression of CSF2RB was observed in NK cells, T lymphocytes or B lymphocytes, and intermediate levels were observed in plasma cells.

GM-CSF stimulation of intestinal immune cells demonstrates broad pSTAT5 induction, but limited pAKT and pERK induction

Previous studies have reported that GM-CSF-deficient mice have impaired mononuclear effector cell function, resulting in reduced regulatory T cell numbers and impaired oral tolerance[21]. In a separate study, administration of neutralizing anti-GM-CSF antibodies to NOD2-deficient, piroxicam-treated mice worsens ileal inflammation[22]. We therefore sought to define the signaling pathways and intestinal immune cells induced by GM-CSF stimulation.

We observed significantly higher fractions of monocytes, macrophages and neutrophils in inflamed compared to non-inflamed tissue (Figure 3A). Expression patterns of CSF2RA by cell types largely mirrored those of CSF2RB, although expression is highly variable in expression between cell subtypes (Figure 3C). We observed greater induction of pSTAT5 compared to pAKT or pERK (Figure 3B). Specifically, within inflamed tissues, we observe GM-CSF stimulated pSTAT5 induction of 2.0- and 3.0-fold (Figure 3B, *P* value = 0.034 and 0.015, respectively) in monocytes and macrophages, respectively, by Cytof, but no significant induction of pAKT or pERK. We corroborated these findings using phosphoflow cytometry in CD14+ intestinal cells, observing a 4.0fold induction (*P* value = 0.031) of pSTAT5, with no induction of pAKT nor pERK observed (Supplementary Figure S2). In contrast to GM-CSF stimulation, PMA-stimulation in CD14+ intestinal cells resulted in significant induction of pAKT and pERK, with 2.0- and 2.6-fold induction, respectively (*P* values = 0.032 and 0.032, Supplementary Figure S2). In our Cytof analyses, cell subtypes demonstrating induction of pSTAT5 with GM-CSF stimulation largely mirror expression patterns of CSF2RA and CSF2RB; comparable induction of pSTAT5 is broadly observed, with a couple of exceptions (Figure 3B, 3D). Plasma cells, despite expressing both receptor subunits, do not demonstrate pSTAT5 induction, and neutrophils demonstrate relatively modest induction. No significant differences in pSTAT5 intensity were observed between uninflamed and inflamed tissue in GM-CSF responsive cells except monocytes (Figure 3D). However, the estimates of pSTAT5 induction are based, in some cases, on ten-fold fewer monocytes (Figure 3A) from non-inflamed compared to inflamed tissues, so estimates of pSTAT5 fold-induction are subject to greater variability.

Loss–of-function, dominant negative effect of CSF2RB frameshift on STAT5 activation in co-transfection studies

Given the previously reported tolerizing effects of GM-CSF in the intestine, it would be predicted that the CD-associated frameshift mutation in *CSF2RB* would result in decreased GM-CSF signaling[21]. Furthermore, the expected number of homozygotes was observed among CD cases, assuming Hardy-Weinberg equilibrium; namely, given the allele frequency

estimate of 3.2% in CD cases, 3.1 *CSF2RB* frameshift homozygotes are predicted, with 4 cases observed. This general concordance is in contrast to the loss-of-function variants in *NOD2*, where a large excess of homozygous carriers are observed in CD cases. A prior report on the crystal structure of the ternary GM-CSF, CSF2RB, and the ligand-specific CSF2RA (colony-stimulating factor receptor, alpha chain) complex revealed an unexpected dodecamer, and possibly higher-order complex, with dodecamer assembly required for activation and signaling[23]. Both the genetic findings and crystal structure suggest a possible dominant negative loss-of-function for the *CSF2RB* frameshift mutation.

To directly assess this, we transfected CSF2RA with wild-type, mutant or equimolar amounts of wild-type and mutant forms of CSF2RB into HEK293 cells and stimulated cells with GM-CSF to assess signaling as measured by pSTAT5 expression (Figure 4A). In five replicate experiments, we observed profoundly decreased pSTAT5 induction (0.16-fold, *P* value = 0.001) with mutant compared to wild-type CSF2RB. Importantly, we also observed a markedly lower pSTAT5 induction (0.38-fold, P value < 0.01) with wild-type plus mutant CSF2RB co-transfection compared to wild-type control (Figure 4B). This would suggest that the mutant form of CSF2RB acts antagonistically to the wild-type allele.

Surface and intracellular levels of CSF2RB were similar for wild-type, mutant and wild-type plus mutant CSF2RB transfectants (Figure 4C), such that the decrease in GM-CSF-induced signaling was not attributable to decreased surface expression of mutant CSF2RB. Moreover, wild-type and mutant CSF2RB colocalized at the plasma membrane upon GM-CSF stimulation (Figure 4D). Collectively, these findings are consistent with a dominant negative effect of the *CSF2RB* frameshift mutation on GM-CSF signaling. This would predict that heterozygous carriers, representing slightly over 6% of AJ CD patients, would have impaired signaling with GM-CSF.

Decreased STAT5 activation and aldehyde dehydrogenase induction in monocytes from CSF2RB frameshift carriers

We next examined the functional effects of GM-CSF signaling in peripheral blood monocytes in wild-type and heterozygous carriers of the *CSF2RB* frameshift. Monocytes were serum starved overnight, then stimulated with 10, 2 and 0.4 ng/ml of GM-CSF for 15 minutes. Decreased pSTAT5 levels were observed in frameshift compared to wild-type monocytes (Figure 5A–B). In contrast, while robust induction of pERK and pAKT was observed with GM-CSF stimulation, no genotype-dependent differences were observed (Figure 5C–D).

GM-CSF modulates mononuclear cell phagocyte differentiation broadly and its tolerizing effects within the intestine are mediated by retinoic acid[21]. Retinoic acid is produced from its precursor vitamin A (retinol) via the sequential actions of retinol dehydrogenase and aldehyde dehydrogenase enzymes. Aldehyde dehydrogenase activity is mediated via multiple gene family members, and enzymatic activity was measured using the flow cytometry-based Aldefluor assay. Significantly greater Aldehyde dehydrogenase activity was observed in wild-type compared to *CSF2RB* frameshift monocyte-derived macrophages (Figure 5E–F), with highly consistent results observed in three paired replicates (Figure 5F).

The central role of altered STAT5 activation in mediating pathogenic effects of the *CSF2RB* frameshift mutation is highlighted by both its predominant effect in intestinal signaling (Figure 3B) compared to pAKT and pERK, as well as the absence of genotype-dependent differences for the latter two pathways (Figure 5D). Prior studies have shown that GM-CSF signaling via dodecamer complex assembly enhances JAK2 phosphorylation, which in turn phosphorylates multiple tyrosine residues of the CSF2RB cytoplasmic domain (Figure 6A–B) [23, 24]. More proximal tyrosine residues (Tyr577, Tyr612, and Tyr695) of CSF2RB mediate MAP kinase[24] and PI3kinase[25] recruitment and activation, whereas more distal residues (Tyr612, Tyr695, Tyr750, and Tyr806) mediate STAT5 activation[24], although functional overlap and redundancies have been reported[26]. Taken together, the present findings indicate that the pathogenic effects of the CD-associated *CSF2RB* frameshift mutation, which occurs at codon 708, resulting in truncation at codon 729, are significantly driven by impaired pSTAT5 activation.

Conclusions

In this study, we have surveyed the role of uncommon frameshift mutations in Ashkenazi Jewish CD. For most common variants reported in GWAS, the genetic architecture among Ashkenazim is similar to non-Jewish European ancestry populations, with similar directions of risk alleles observed for the large majority of loci[19]. Rare variant analyses focused on Jewish populations may be particularly fruitful because of the higher CD prevalence and unique population history of the Ashkenazim. While the *CSF2RB* variant is most common in AJ populations, the similar trends and odds ratios observed in non-AJ populations indicates a more general role of impaired GM-CSF signaling in CD pathogenesis. Inferences on AJ population history based on whole genome sequencing suggest a reduction of population size, with a severe population bottleneck to as few as 350 individuals as recently as 28 thousand years ago[27]. This small founder population enables cataloguing of the vast majority of pre-bottleneck AJ variation by sequencing a relatively modest number of individuals. This history would also suggest an increased power to detect variants of larger effect that drifted to higher frequencies during the bottleneck[27]. Sequencing 47 CD cases has 0.62, 0.77 and 0.86 probabilities of observing variants of 1.0%, 1.5% and 2.0% frequency, respectively (Supplemental table 1), suggesting that we have tested a large fraction of the rare variants for which presently collected case-control cohorts are powered to detect association. Sequencing larger numbers of cases will identify successively rarer variants, some of which may have larger effect sizes, but will also require much larger cohort sizes to establish association (Supplemental table 1).

A central value of genetic association studies is in defining primary pathogenic factors driving disease expression. In CD, elevated levels of anti-GM-CSF autoantibodies are correlated with ileal location and worse clinical outcomes[22]. It is often unclear whether select autoantibodies observed in certain immune-mediated diseases are pathogenic, or merely reflect secondary effects of increased immune activity; the correlation of anti-GM-CSF autoantibodies with more severe CD[22] might merely reflect increased general immune activity in such cases. The present findings of a loss-of-function frameshift variant in *CSF2RB* argue for a primary pathogenic role of decreased GM-CSF signaling in driving a subset of CD cases. As many as 30% of CD cases have elevated levels of anti-GM-CSF

antibodies, which can neutralize GM-CSF activity[22], indicating that impaired GM-CSF signaling plays a major role in CD. These findings are consistent with a failure of the acute inflammatory response[28]. Pulmonary alveolar proteinosis is also associated with elevated auto-antibodies to GM-CSF, albeit at much higher titers than those observed in CD [22, 29]. Of interest, a case of adult-onset pulmonary alveolar proteinosis was reported in an individual lacking anti-GMF-CSF autoantibodies, but instead carrying an autosomal recessive, single base pair deletion at nucleotide 631 of *CSF2RB*. This more proximal mutation was associated with the absence of surface expression of CSF2RB and the complete absence of GM-CSF induced signaling, in contrast to the hypomorphic effects of the *CSF2RB* mutation described here. Taken together, these findings highlight phenotypic variability resulting from variable degrees of impaired GM-CSF signaling. In addition, analyses of CSF2RA (CD116) expression in circulating granulocytes and monocytes demonstrated decreased expression of CD116 as a distinguishing feature of IBD, broadly implicating impaired GM-CSF signaling in disease pathogenesis[30].

Given the essential role of GM-CSF in maintaining intestinal tolerance[21], the present association of a frameshift variant attenuating GM-CSF signaling and decreasing aldehyde dehydrogenase activity sharply contrasts CD pathogenesis from rheumatoid arthritis, where anti-GM-CSF monoclonal antibodies have demonstrated efficacy[31]. While a meta-analysis of three studies involving GM-CSF administration did not establish superiority of treatment in CD, all studies showed a trend toward efficacy[32]; it is possible that future studies may identify subsets of CD patients for which agonism of the GM-CSF pathway is most effective.

The correlation of GM-CSF autoantibodies in complicated CD(**REF) strongly implicates decreased GM-CSF signaling in the pathogenesis of the *CSF2RB* frameshift mutation. However, given the high expression of CSF2RB on intestinal eosinophils (Figure 2A), future studies exploring the comparative effects of IL-3 and IL-5 on altered eosinophil differentiation and function would provide additional insight. In addition, because GM-CSF-deficient mice demonstrated impaired epithelial proliferation and ulcer healing in response to DSS-induced injury(PMID: 23325885), future studies examining the effects of the *CSF2RB* mutation on altered epithelial cell function would potentially highlight additional contributing disease mechanisms.

Given the prior associations of loss-of-function risk alleles involving the *NOD2*[4] and autophagy innate immune pathways[33, 34] the present identification of a loss-of-function *CSF2RB* frameshift mutation highlights the multiple innate immune deficiencies associated with CD. In further support of this, patients with chronic granulomatous disease, characterized by phagocytic defects, can present with intestinal inflammation[35]. While the present major therapies involve general immunosuppression, if CD is, in a subset of patients, a primary immunodeficiency of macrophages[36], a fundamental paradigm shift may be in order; enhancing specific phagocyte cell subset function may be a more effective means of developing the novel therapies needed in CD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Ling-Shiang Chuang[1,†], Nicole Villaverde[1,†], Ken Y. Hui[20,2,†], Arthur Mortha[34], Adeeb Rahman[1], Adam P. Levine[44], Talin Haritunians[3], Sok Meng Evelyn Ng[20], Wei Zhang[20], Nai-Yun Hsu[1], Jody-Ann Facey[1], Tramy Luong[1], Heriberto Fernandez-Hernandez[1], Dalin Li[3], Manuel Rivas[38,39,40], Elena R. Schiff[44], Alexander Gusev[6], L. Phillip Schumm[7], Beatrice M. Bowen[20,4], Yashoda Sharma[20,1], Kaida Ning[20,41], Romain Remark[34], Sacha Gnjatic[35], Peter Legnani[36], James George[36], Bruce E. Sands[36], Joanne M. Stempak[9,10], Lisa W. Datta[11,12], Seth Lipka[13], Seymour Katz[16], Adam S. Cheifetz[17], Nir Barzilai[19], Nikolas Pontikos[44], Clara Abraham[20], Marla J. Dubinsky[36,46], Stephan Targan[3], Kent Taylor[47], Jerome I Rotter[47], Ellen J. Scherl[48], Robert J. Desnick[1], Maria T. Abreu[21], Hongyu Zhao[22], Gil Atzmon[19], Itsik Pe'er[8], Subra Kugathasan[23], Hakon Hakonarson[24,25], Jacob L. McCauley[26,27], Todd Lencz[28], Ariel Darvasi[29], Vincent Plagnol[45], Mark S. Silverberg[9,10], Aleixo M. Muise[30,31], Steven R. Brant[11,12], Mark J. Daly[38,42,43], Anthony W. Segal[44], Richard H. Duerr[32], Miriam Merad[34], Dermot P.B. McGovern[3], Inga Peter[1,†], and Judy H. Cho[1,36,52,†,*]

Affiliations

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[2]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA 06520

[3]F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA 90048

[4]Department of Genetics, Yale University, New Haven, CT, USA 06520

[5]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea

[6]Department of Epidemiology, Harvard University, Boston, MA, USA 02115

[7]Department of Health Studies, University of Chicago, Chicago, IL, USA 60637

[8]Department of Computer Science, Columbia University, New York, NY, USA 10027

[9]Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital, Toronto, Ontario, Canada M5T3L9

[10]Department of Medicine, University of Toronto, Toronto, Ontario, Canada M5G1X5

[11]Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA 21231

[12]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA 21231

[13]Department of Internal Medicine, University of South Florida, Tampa, FL, USA 33606

[14]Montreal Heart Institute, Montréal, QC, Canada H1T1C8

[15]Université de Montréal, Montréal, QC, Canada H1T1C8

[16]Department of Medicine, New York University School of Medicine, New York, NY, USA10016

[17]Division of Gastroenterology, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215

[18]Jill Roberts Center for Inflammatory Bowel Disease, Weill Medical College of Cornell University, New York, NY, USA 10021

[19]Department of Medicine, Albert Einstein College of Medicine, Bronx, NY, USA 10461

[20]Section of Digestive Diseases, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA 06520

[21]Division of Gastroenterology, University of Miami, Miller School of Medicine, Miami, Florida, USA 33136

[22]Department of Biostatistics, Yale University, New Haven, CT, USA 06520

[23]Department of Pediatrics, Emory University, Atlanta, GA, USA 30322

[24]Centre for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA 19104

[25]Division of Human Genetics, University of Pennsylvania, Philadelphia, PA, USA 19014

[26]John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, Florida, USA 33136

[27]Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, Florida, USA 33136

[28]Feinstein Institute for Medical Research, North Shore – Long Island Jewish Health System, Manhasset, NY, USA 11030

[29]Department of Genetics, The Hebrew University of Jerusalem, Jerusalem, Israel 91904

[30]Inflammatory Bowel Disease Centre and Cell Biology Program, The Hospital for Sick Children, Toronto, ON, Canada M5G1X8

[31]Department of Pediatrics, University of Toronto, Toronto, ON, Canada M5G1X8

[32]Division of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA, 15261

[33]Section of Gastroenterology and Hepatology, Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA 06520

[34]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[35]Division of Hematology and Medical Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[36]Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[37]Department of Anesthesiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[38]Department of Medical and Population Genetics, Broad Institute, Cambridge, MA, USA 02142

[39]Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA 02114

[40]Nuffield Department of Clinical Medicine, Wellcome Trust Centre for Human Genetics Research, University of Oxford, Oxford, UK OX3 UBN

[41]Department of Molecular and Computational Biology, University of Southern California, Los Angeles, CA, USA 90033

[42]Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital, Boston, MA USA 02114

[43]Department of Genetics, Harvard Medical School, Boston, MA, USA 02115

[44]Centre for Molecular Medicine, Division of Medicine, University College, London, UK WC1E 6JF

[45]Genetics Institute, Division of Biosciences, University College, London, UK WC1E 6BT

[46]Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[47]Institute for Translational Genomics and Population Sciences, Division of Genomic Outcomes, Harbor-UCLA Medical Center, Torrance, CA, USA 90502

[48]The Division of Gastroenterology & Hepatology, Sanford I. Weill College of Cornell University—New York Presbyterian Hospital, New York, NY, USA 10021

[49]Department of Medicine, New York University School of Medicine, New York, NY, USA 10016

[50]St. Francis Hospital, Roslyn, New York, NY USA 11021

[51]Department of Genetics, Graduate School of Public Health; University of Pittsburgh, PA, USA 15261

[52]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

Acknowledgments

We thank the patients that participated in this study and the staff that assisted in their recruitment. We appreciate the sample contribution of The Charles Bronfman Institute for Personalized Medicine's Biobank at Mount Sinai. The Flow Cytometry CORE at the Icahn School of Medicine at Mount Sinai provided training, consultation, and technical assistance for CyTOF and flow cytometry. Microscopy was performed at the Microscopy CORE at the Icahn School of Medicine at Mount Sinai.

Funding source: The NIH (DK092235), Inflammatory Bowel Disease Genetics Consortium (DK062429), Genetic Research Center at the Icahn School of Medicine (DK062422), New York Crohn's Foundation, Consortium ancillary RO1 (DK099097), U01 (DK062431), Inflammatory Bowel Disease Genetic Research Chair, RO1 (DK062420), RO1 (CA141743), the Atran Foundation and the Sanford J. Grossman Charitable Trust financially supported this work. Researchers at UCL were funded by the Wellcome Trust, Charles Wolfson Charitable Trust, and the Irwin Joffe Memorial Fellowship. IBD Research at Cedars-Sinai is supported by USPHS grant PO1 (DK046763) and the Cedars-Sinai F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute Research Funds. Project investigators are supported by The Helmsley Charitable Trust (D.P.B.M.), The European Union (D.P.B.M.), The Crohn's and Colitis Foundation of America (CCFA) (D.P.B.M.), The Joshua L. and Lisa Z. Greer Chair in IBD Genetics (D.P.B.M.), and grants, DK062413, DK046763-19, AI067068, HS021747 (D.P.B.M.).

References and Notes

1. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015
2. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012; 491:119–124. [PubMed: 23128233]
3. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with.
4. Ogura Y, Bonen DK, Inohara N, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature. 2001; 411:603–606. [PubMed: 11385577]
5. Li J, Moran T, Swanson E, et al. Regulation of IL-8 and IL-1beta expression in Crohn's disease associated NOD2/CARD15 mutations. Hum Mol Genet. 2004; 13:1715–1725. [PubMed: 15198989]
6. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014; 508:469–476. [PubMed: 24759409]
7. Bernstein CN, Rawsthorne P, Cheang M, et al. A population-based case control study of potential risk factors for IBD. Am J Gastroenterol. 2006; 101:993–1002. [PubMed: 16696783]
8. Mayberry JF, Judd D, Smart H, et al. Crohn's disease in Jewish people--an epidemiological study in south-east Wales. Digestion. 1986; 35:237–240. [PubMed: 3817333]
9. Ostrer H, Skorecki K. The population genetics of the Jewish people. Hum Genet. 2013; 132:119–127. [PubMed: 23052947]
10. Woodage T, King SM, Wacholder S, et al. The APCI1307K allele and cancer risk in a community-based study of Ashkenazi Jews. Nat Genet. 1998; 20:62–65. [PubMed: 9731533]
11. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell. 1996; 87:159–170. [PubMed: 8861899]
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]
13. Hinrichs AS, Karolchik D, Baertsch R, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006; 34:D590–D598. [PubMed: 16381938]
14. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]

15. Grove ML, Yu B, Cochran BJ, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. PLoS One. 2013; 8:e68095. [PubMed: 23874508]
16. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
17. Chen TJ, Kotecha N. Cytobank: providing an analytics platform for community cytometry data analysis and collaboration. Curr Top Microbiol Immunol. 2014; 377:127–157. [PubMed: 24590675]
18. Qiu P, Simonds EF, Bendall SC, et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat Biotechnol. 2011; 29:886–891. [PubMed: 21964415]
19. Kenny EE, Pe'er I, Karban A, et al. A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. PLoS Genet. 2012; 8:e1002559. [PubMed: 22412388]
20. Bain CC, Scott CL, Uronen-Hansson H, et al. Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6Chi monocyte precursors. Mucosal Immunol. 2013; 6:498–510. [PubMed: 22990622]
21. Mortha A, Chudnovskiy A, Hashimoto D, et al. Microbiota-dependent crosstalk between macrophages and ILC3 promotes intestinal homeostasis. Science. 2014; 343:1249288. [PubMed: 24625929]
22. Han X, Uchida K, Jurickova I, et al. Granulocyte-macrophage colony-stimulating factor autoantibodies in murine ileitis and progressive ileal Crohn's disease. Gastroenterology. 2009; 136:1261–1271. e1–3. [PubMed: 19230854]
23. Hansen G, Hercus TR, McClure BJ, et al. The structure of the GM-CSF receptor complex reveals a distinct mode of cytokine receptor activation. Cell. 2008; 134:496–507. [PubMed: 18692472]
24. Itoh T, Liu R, Yokota T, et al. Definition of the role of tyrosine residues of the common beta subunit regulating multiple signaling pathways of granulocyte-macrophage colony-stimulating factor receptor. Mol Cell Biol. 1998; 18:742–752. [PubMed: 9447970]
25. Dijkers PF, van Dijk TB, de Groot RP, et al. Regulation and function of protein kinase B and MAP kinase activation by the IL-5/GM-CSF/IL-3 receptor. Oncogene. 1999; 18:3334–3342. [PubMed: 10362354]
26. de Groot RP, Coffer PJ, Koenderman L. Regulation of proliferation, differentiation and survival by the IL-3/IL-5/GM-CSF receptor family. Cell Signal. 1998; 10:619–628. [PubMed: 9794243]
27. Carmi S, Hui KY, Kochav E, et al. Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. Nat Commun. 2014; 5:4835. [PubMed: 25203624]
28. Smith AM, Rahman FZ, Hayee B, et al. Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn's disease. J Exp Med. 2009; 206:1883–1897. [PubMed: 19652016]
29. Uchida K, Beck DC, Yamamoto T, et al. GM-CSF autoantibodies and neutrophil dysfunction in pulmonary alveolar proteinosis. N Engl J Med. 2007; 356:567–579. [PubMed: 17287477]
30. Goldstein JI, Kominsky DJ, Jacobson N, et al. Defective leukocyte GM-CSF receptor (CD116) expression and function in inflammatory bowel disease. Gastroenterology. 2011; 141:208–216. [PubMed: 21557945]
31. Behrens F, Tak PP, Ostergaard M, et al. MOR103, a human monoclonal antibody to granulocyte-macrophage colony-stimulating factor, in the treatment of patients with moderate rheumatoid arthritis: results of a phase Ib/IIa randomised, double-blind, placebo-controlled, dose-escalation trial. Ann Rheum Dis. 2015; 74:1058–1064. [PubMed: 24534756]
32. Roth L, MacDonald JK, McDonald JW, et al. Sargramostim (GM-CSF) for induction of remission in Crohn's disease: a cochrane inflammatory bowel disease and functional bowel disorders systematic review of randomized trials. Inflamm Bowel Dis. 2012; 18:1333–1339. [PubMed: 22552871]
33. Murthy A, Li Y, Peng I, et al. A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. Nature. 2014; 506:456–462. [PubMed: 24553140]
34. Lassen KG, Kuballa P, Conway KL, et al. Atg16L1 T300A variant decreases selective autophagy resulting in altered cytokine signaling and decreased antibacterial defense. Proc Natl Acad Sci U S A. 2014; 111:7741–7746. [PubMed: 24821797]

35. Marciano BE, Rosenzweig SD, Kleiner DE, et al. Gastrointestinal involvement in chronic granulomatous disease. Pediatrics. 2004; 114:462–468. [PubMed: 15286231]

36. Casanova JL, Abel L. Revisiting Crohn's disease as a primary immunodeficiency of macrophages. J Exp Med. 2009; 206:1839–1843. [PubMed: 19687225]

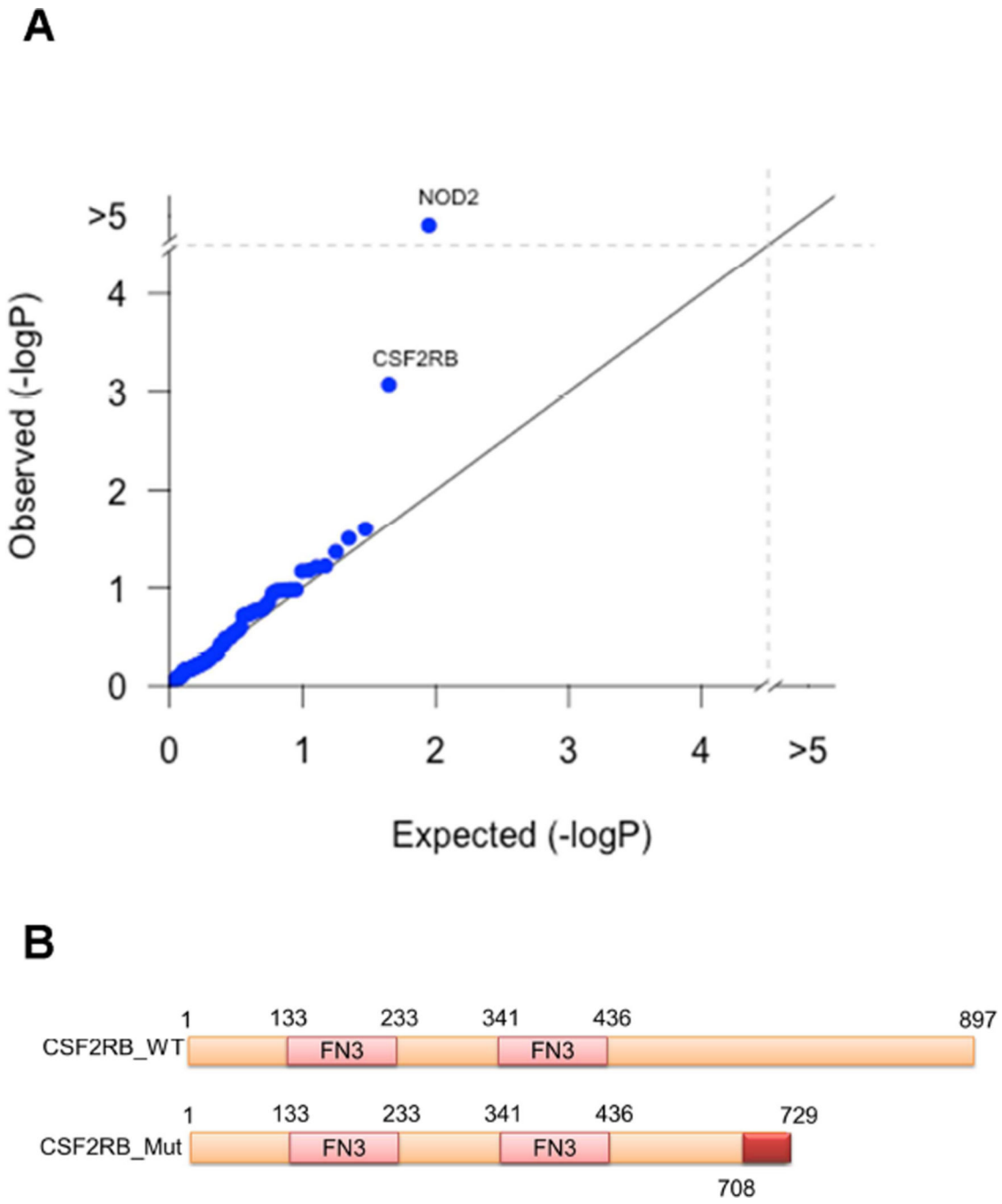

Figure 1. Frameshift mutation in *CSF2RB* associated to CD

A) Quantile-quantile plot of 88 frameshift mutations in the CD case-control discovery cohort (1477 cases, 2614 controls). **B)** Schematic of wild-type (WT) and frameshift-containing (Mut) *CSF2RB* transcript, with fibronectin type III (FN3) domains indicated in pink, and altered frameshift region prior to stop codon indicated in red.

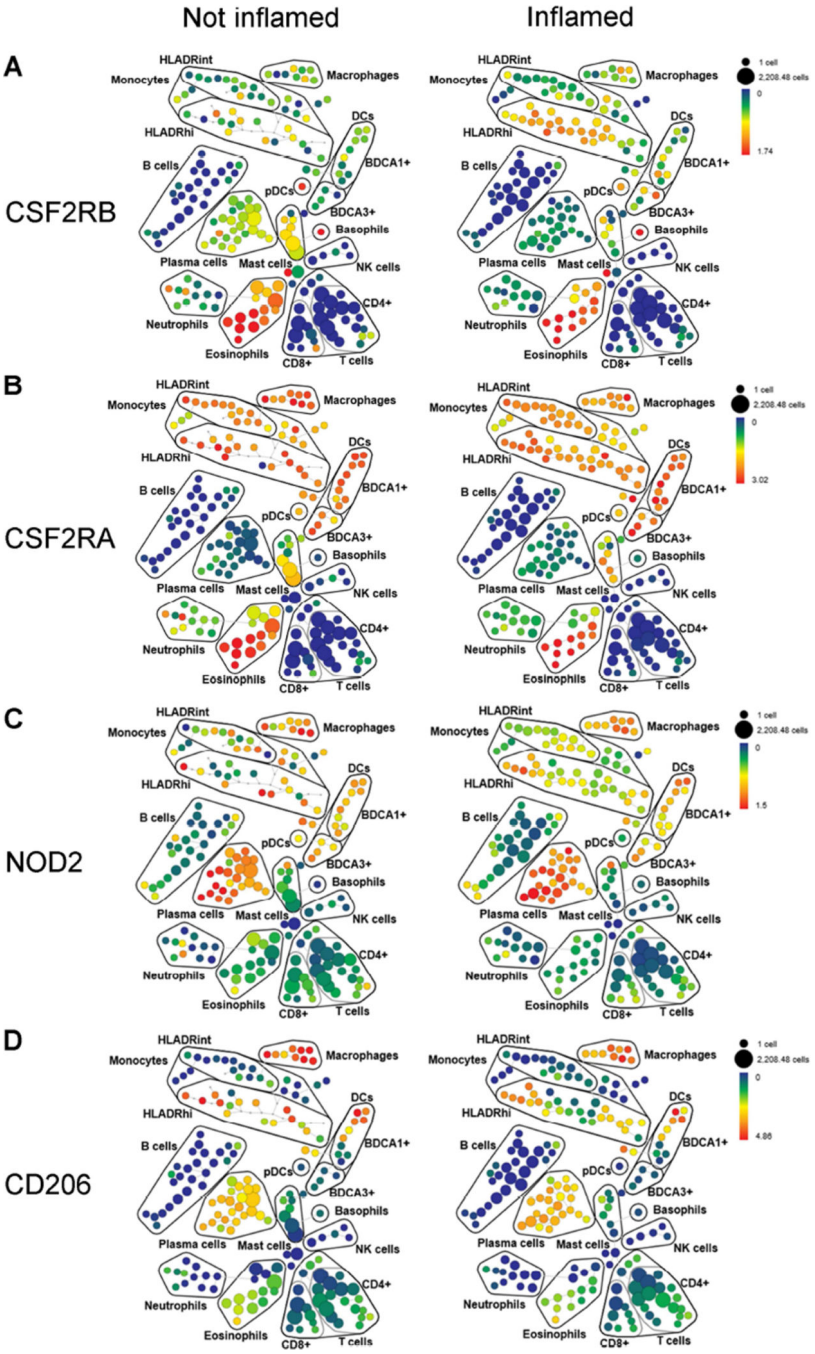

Figure 2. CyTOF analysis of cellular expression of CSF2RB, CSF2RA, NOD2 and CD206 in CD intestine

Representative cluster plot of CyTOF output for not inflamed (left) and inflamed (right) regions from resected terminal ileum from a CD patient. Colors correspond to the average quantity of CSF2RB **(A)**, CSF2RA **(B)**, NOD2 **(C)** and CD206 **(D)** detected per cell (blue to red continuum corresponding with low to high expression). The size of each node corresponds to the number of individual cells detected.

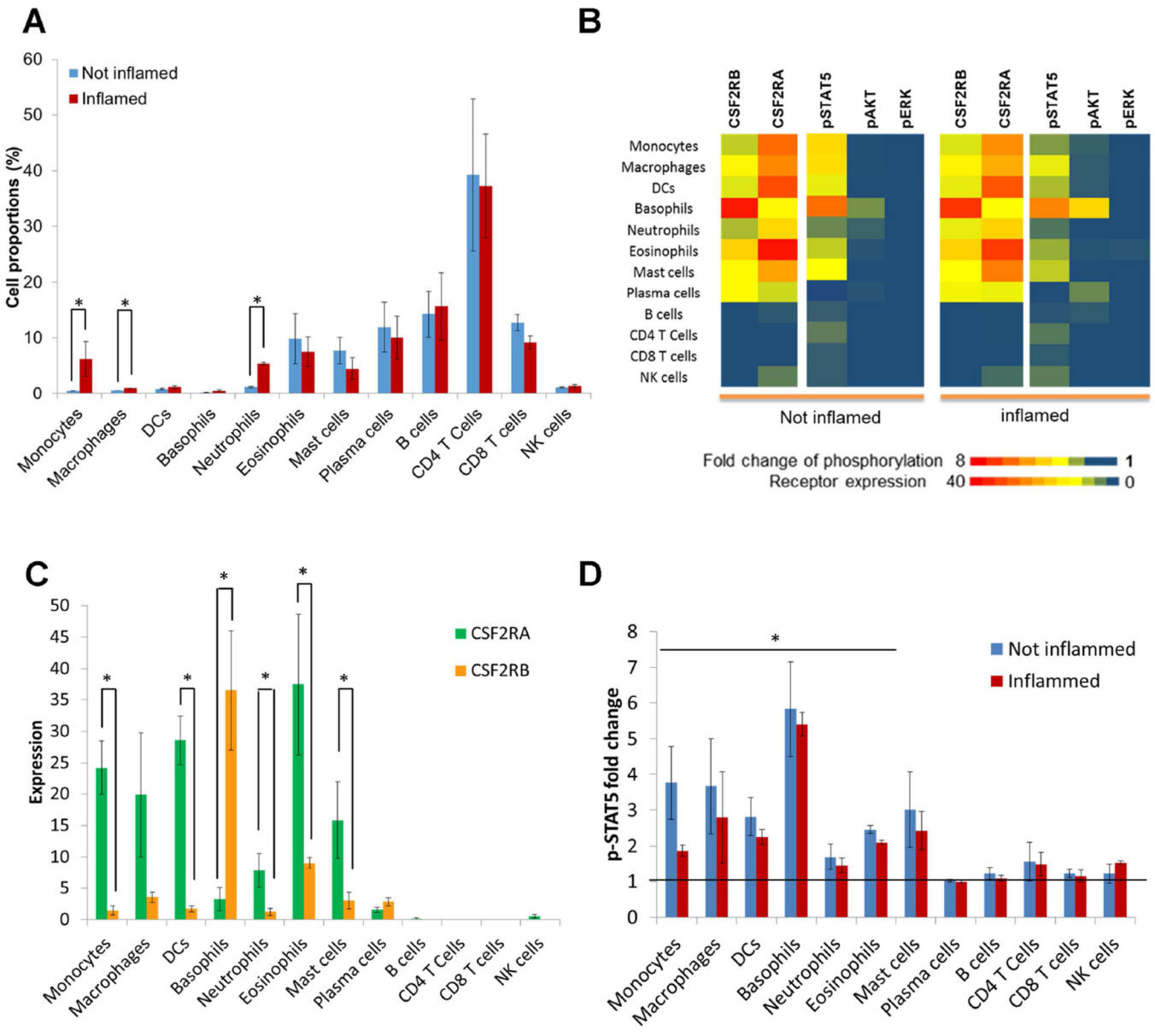

Figure 3. CyTOF analysis of intestinal cells from CD patients defines GM-CSF responding cell subsets

(A) Summary of proportions of immune cell subtypes detected in not-inflamed and inflamed ileal specimens (n=3). **(B)** Relative quantities of receptor subunits (left) and intracellular signaling molecules (right) by immune cell subtype in both inflamed (right) (n=3) and not inflamed (left) regions of resected terminal ileum. **(C)** Quantification of CSF2RA (green) and CSF2RB (orange) expression in immune cell subsets. **(D)** Fold-change of phosphorylated STAT5 in immune cell subsets in response to GM-CSF treatment in inflamed (red) and not inflamed (blue) ileal samples. Significance estimates in (D) are compared to no GM-CSF treatment. *, $P < .05$.

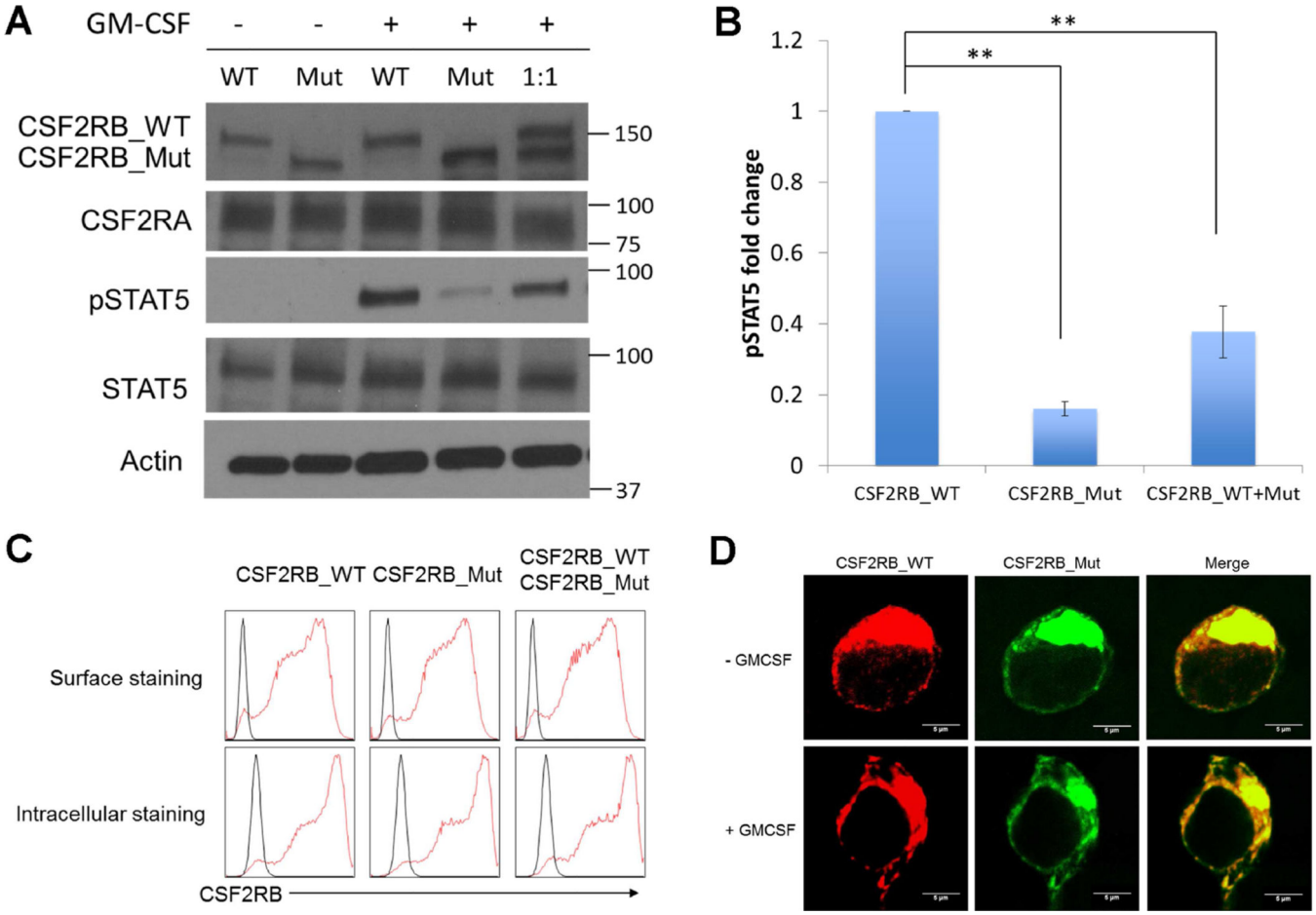

Figure 4. The CD-associated frameshift in *CSF2RB* results in a truncated protein and attenuates GM-CSF-induced STAT5 phosphorylation in a dominant negative manner

(A) HEK293 cells were transfected with CSF2RA + wild-type (WT) and/or frameshift mutation (Mut) CSF2RB, alone or in combination. Cells were then treated with 10ng/ml GM-CSF for 15min. Representative western blot displaying products of transfection and phospho-STAT5. STAT5 and β-actin are shown as loading controls. (**B**) Quantitative levels of pSTAT5 relative to wild-type control (CSF2RB_WT) from five replicated experiments. **, $P < .01$. (**C**) Transfected HEK293 cells were stained with isotype antibodies (black line) or anti-CSF2RB antibodies (red line) on their surface (top row) and intracellularly (bottom row). Representative histograms are shown. (**D**) Confocal fluorescent micrograph illustrating intracellular localization of wild-type (left), mutant (middle) construct products and merge (right) in cells co-transfected with both constructs with 10ng/ml GM-CSF treatment for 15min. Yellow areas indicate the co-localization of CSF2RB_WT and CSF2RB_Mut receptors. Scale bar: 5μm.

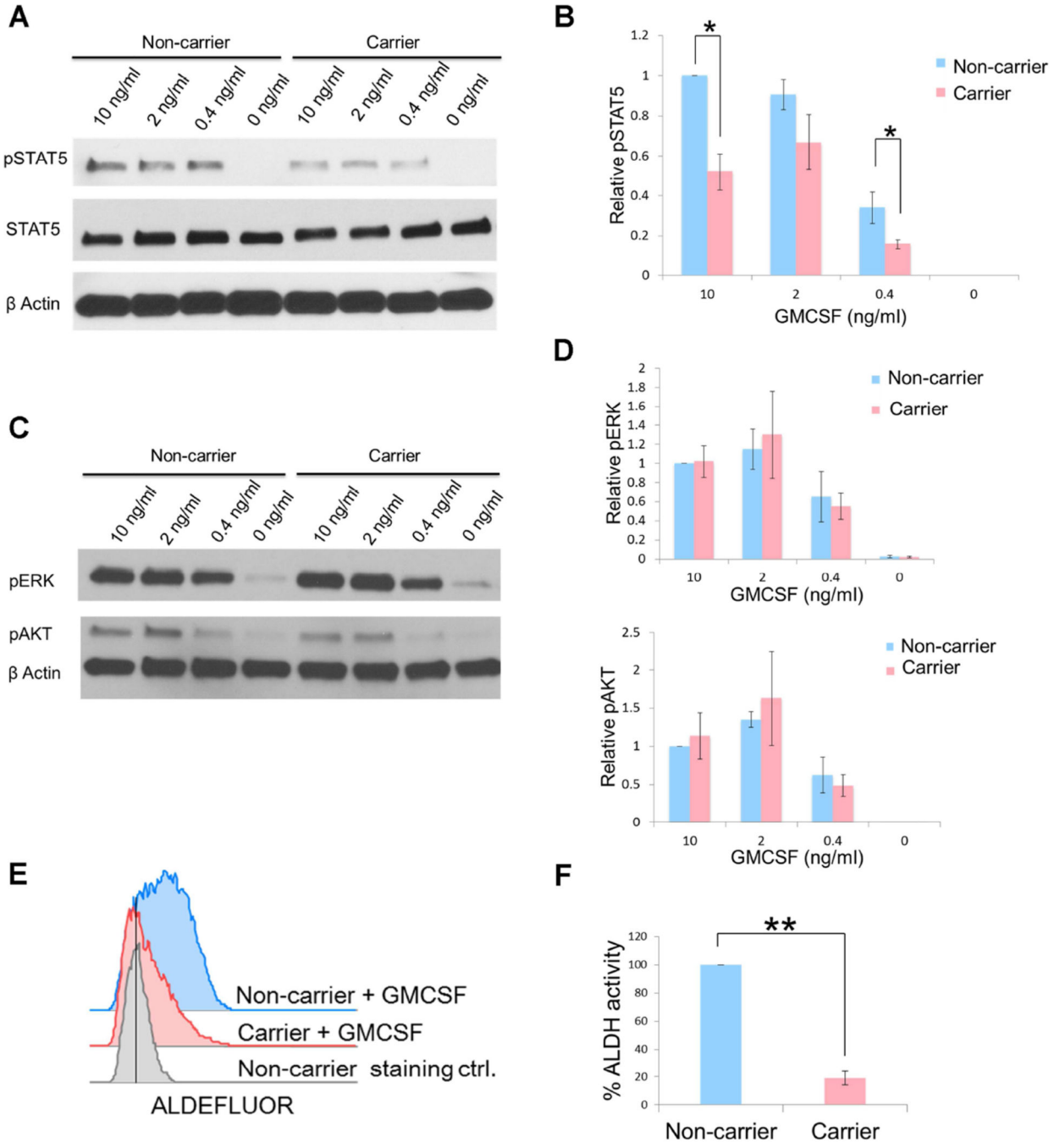

Figure 5. Primary monocytes from *CSF2RB* frameshift carriers demonstrate diminished GM-CSF mediated pSTAT5 and aldehyde dehydrogenase activity with intact ERK and AKT activation

(A) Representative immunoblot of peripheral monocytes from a *CSF2RB* frameshift carrier and matched control displaying pSTAT5 and total STAT5 15 min after stimulation with various concentrations of GM-CSF. (**B**) Quantification of pSTAT5 levels after GM-CSF stimulation in three phenotypically matched pairs with normalizing to loading control. *, $P < .05$. (**C**) Immunoblot of pERK and pAKT in a frameshift carrier and a matched non-carrier after GM-CSF stimulation. (**D**) Quantification of pERK and pAKT in 3 matched pairs of

frameshift carriers and non-carriers. (**E**) Activity of aldehyde dehydrogenase (ALDH), an enzymatic catalyst involved in retinoic acid synthesis, as measured by flow cytometry detecting ALDEFLUOR after 10ng/ml of GM-CSF stimulation for 15 minutes, in a *CSF2RB* frameshift non-carrier (top) and a matched carrier (middle). (**F**) Quantification of ALDH activity by mean ALDEFLUOR intensity in 3 pairs of frameshift carriers and non-carriers. **, $P < .01$.

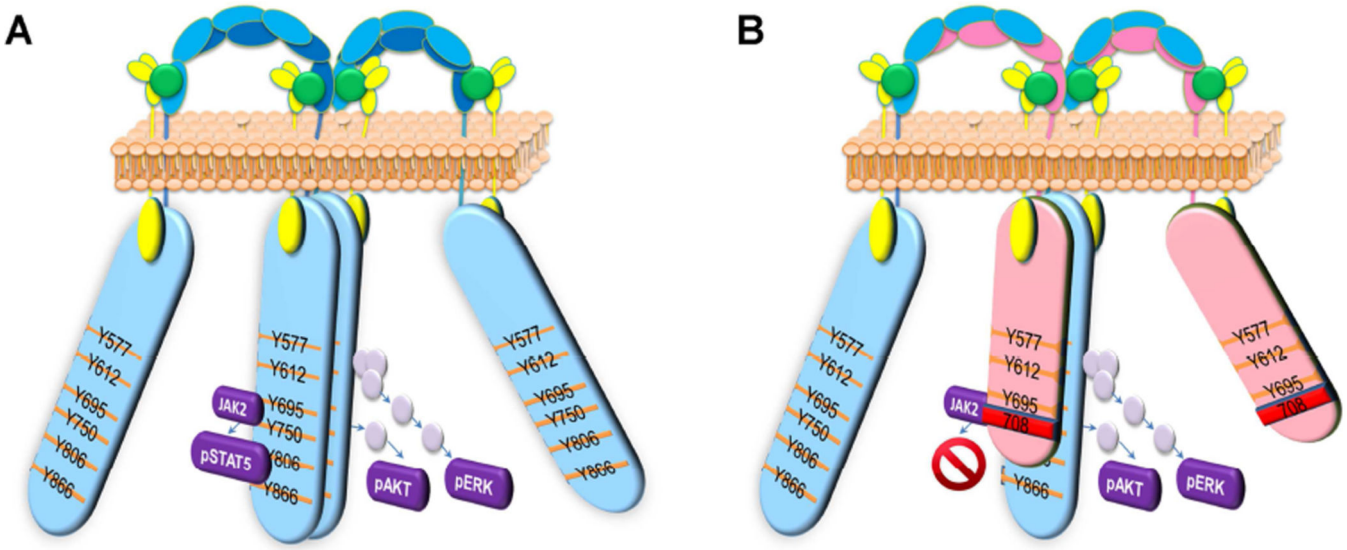

Figure 6. Schematic model of altered GM-CSF signaling with the *CSF2RB* frameshift mutation

GM-CSF signaling includes GM-CSF cytokine (green), CSF2RA (yellow) and CSF2RB (wild-type in blue, mutant in pink) components. Crystal structure analysis reveals a high affinity, hexameric bridge. JAK2, associated with the CSF2RB cytoplasmic domain, is able to dimerize and transphosphorylate in the dodecamer, but not in the hexamer complex. With GM-CSF stimulation of wild-type CSF2RB (A, blue), potent induction of pSTAT5, pAKT, and pERK is observed. However, with heterozygous *CSF2RB* frameshift carriage (wild-type in blue, mutant in pink CSF2RB with the frameshift region in red) pAKT and pERK signaling is largely intact, reflecting interaction with more proximal tyrosine residues within the CSF2RB tail. In contrast, JAK2 interacts with more distal tyrosine residues within CSF2RB. Given the JAK2 requirement for dimerization and transphophorylation in the dodecamer complex, the *CSF2RB* frameshift mutation would impair wild-type *CSF2RB* signaling, consistent with results observed in both cellular transfectant (Figure 4) and primary monocyte (Figure 5) studies.

Table 1

Association results for the *CSF2RB* frameshift mutation

	N_{CD}	N_{ctrl}	MAF_{CD}	MAF_{ctrl}	*P* value	OR
AJ Discovery cohort	1477	2614	0.032	0.020	8.52×10^{-4}	1.6
AJ Replication cohort	1515	7052	0.033	0.022	5.39×10^{-4}	1.5
Total	2992	9666	0.032	0.022	3.42×10^{-6}	1.5

MAF, minor allele frequency; OR, odds ratio