Emory University School of Medicine
Department of Pediatrics
2015 Uppergate Drive
Atlanta, Georgia 30322

Research Protocol

Title: Risk Stratification and Identification of Immunogenetic and Microbial Markers of Rapid Disease Progression in Children with Crohn's Disease

Investigator: Subra Kugathasan, MD

A. PURPOSE OF THE STUDY

This is an observational prospective cohort study. The purpose of this proposal is to develop and validate a risk stratification model for predicting severe and complicated disease phenotype in children with Crohn's disease (CD). Demographic, clinical data together with immune, genetic and microbial (enteric flora) bio-markers in children with newly diagnosed CD will be collected. Time to develop complications (stricturing or perforating) in these patients will be documented prospectively. Identification of the bio-markers together with prospective clinical follow up will help to identify the estimated 20% of patients who develop severe complicated phenotype within the minimum of 36 months from the diagnosis.

B. HYPOTHESIS / SPECIFIC AIMS

Rationale & Hypothesis:

The currently accepted pathogenic hypothesis for inflammatory bowel disease proposes the role of host genetics, specifically "pre-programmed" host immune responses, and environmental factors such as microbial triggers, including in particular, enteric flora, resulting in disease susceptibility, development, and eventual expression. These factors (host immune/inflammatory cells, intestinal epithelia and microbial flora) and their interactions may also be important determinants of disease phenotype and disease progression. **Therefore, we hypothesize** that *there are identifiable immunologic, genetic, enteric flora profiles along with clinical risk factors that influence the disease onset and phenotype, as well as the natural history of pediatric CD.*

Specific Aims:

Aim 1: To identify demographic, clinical, microbial (i.e., fecal stream characterization), genetic, and/or immunologic risk factors that influence the likelihood of the rapid development of complicated disease phenotypes manifested as, penetrating disease, stricturing disease and need for surgery in children with newly diagnosed CD. The newly formed Pediatric IBD North American Collaborative Research Group (PIBD NACRG) will utilize the existing participating centers in the four current consortia and recruit additional centers to enroll newly diagnosed uncomplicated (nonpenetrating-nonstricturing) pediatric CD patients.

1a: Recruit a minimum of 2000 children (age ≤ 16 years) with newly diagnosed or suspected Crohn's disease using standardized diagnostic criteria by collaborating with participating members of the Pediatric IBD North American Collaborative Research Group. Inflammatory bowel disease (IBD) encompasses both Crohn's disease and ulcerative colitis. However, there are no gold standards for the diagnosis of Crohn's disease despite the diagnostic criteria we use. Some of the children suspected to have Crohn's disease may ultimately be diagnosed with ulcerative colitis or no IBD at all. We will continue to collect follow-up data on those children when the follow-up data is available.
1b: Collect clinical and demographic information from well-characterized newly diagnosed IBD patients.
1c: Collect serum samples at diagnosis and annually for immune response testing for both current serological immune markers including ANCA, ASCA, OMPC, I2 and CBir1 and any newly identified markers, including GM-CSF auto-antibody (GM-CSF Ab).
1d: Collect and genotype samples for known candidate, novel genetic and any newly discovered variants by high throughput SNP technology. This will include the *NOD2, ATG16L1, IRGM,* and *TNFRSF6B* SNPs associated with disease susceptibility and behavior. Since the DNA methylation can change with age, nutritional status, and medication exposure, we will collect DNA annually during the 3 year follow-up period as well.
1e: Evaluate fresh fecal specimens by molecular methodologies to determine relative proportions of enteric microflora in children with newly diagnosed Crohn's disease.
1f: Collect mucosal biopsies in a subset inception cohort (n=200) and stored for future gene expression analysis and specific bacterial species.
1g: Establish a Centralized Data Coordinating Center for data collection, data quality and detailed data analyses (Clinipace) and a Biospecimen Procurement Center and Repository (Emory University) to facilitate subsequent specimen analysis upon acquisition of new knowledge regarding "risk" variables.

Aim 2: Develop and validate risk stratification by stratifying patients into different levels of risk at diagnosis based on clinical, demographic, host microbial ecology, immune and genetic determinants identified in Aim 1. The risk stratification will be completed by state-of-the-art analysis using logistic regression and mathematical modeling among the different risk groups of pediatric CD patients followed over a minimum of 36 month period. Time to complication as defined by time to stricture or time to internal penetration will be included in the analysis.

C. BACKGROUND, SIGNIFICANCE, AND RATIONALE

Emerging data suggests that distinct phenotypes of CD may be the result of the complex interplay between host genetics, immune response and environmental triggers. Our data suggest that approximately 20 percent of patients with CD develop rapidly progressing complicated disease within 3 years of diagnosis. Thus, well-designed, multi-disciplinary investigations of genetic, immune and microbial etiologies of CD, with sufficient power to detect differences in disease phenotype to allow risk stratification and prognostication, are critically needed.

Crohn's Disease Clinical Phenotypes: The observed clinical heterogeneity in CD has led to the development of certain CD classification systems in order to help characterize patients as specific disease phenotypes (i.e. attempting to classify patients into more homogeneous subgroups) based on disease location. However, these classification systems are not ideal.

CD phenotype is not always a static phenomenon. Individual patients may present initially with a particular disease behavior or location, yet may present with a changing clinical pattern over the course of disease. It appears that disease location essentially remains stable over time, yet disease behavior evolves, such that after 20 years of follow up, at least 80% of patients with originally non-complicating disease developed a complication, either penetrating or stricturing in nature. These findings suggest that inflammatory or non-complicating disease behavior may not be a stable phenotype, but just a temporary state that evolves in to one of the two complicating disease states over time. The relationship between disease location and disease behavior has been examined and it has been suggested that the site of disease at diagnosis is predictive of clinical disease course, type of complications and prognosis. Small bowel involvement has been shown to be more often associated with stricturing disease while, ileocolonic disease was associated with penetrating disease. Moreover, the frequency of surgery was higher among patients with ileocolonic disease. To date, the influence of disease location on disease behavior and disease progression have received little attention in the pediatric age group, thereby providing additional merit for this proposal.

Genetic Variation: The increased level of concordance between identical twins and the familial risk of IBD provides strong evidence that genetic factors play an important role in the pathogenesis of IBD. Patients with early-onset disease have a significantly higher frequency of affected relatives, suggesting a greater genetic influence on disease development in childhood. Studies suggested that pedigrees with CD diagnosed at an earlier age have greater linkage evidence for IBD1 on Chromosome 16. Subsequently fine mapping of the IBD1 region of interest led to the discovery of the first CD susceptibility gene, *NOD2/CARD15*. The most consistent associations observed were between NOD2 and ileal disease location and fibrostenosing disease behavior. Various other candidate genes including *DLG5*, *MDR1*, *TNF* polymorphisms and most recently *IL23R* have been implicated in susceptibility to CD. *IL23R* has now been studied among pediatric cohorts, supporting the association with CD. Like *NOD2*, these studies will lead to advancements in the understanding of the role of genetics in conferring clinical heterogeneity. With the advancing field of IBD genetics, the search for the genetic contributions and future candidate genes to the various disease phenotypes will need to be intensified.

The identification of the role of autophagy in both the susceptibility to CD and in its etiopathogenesis has opened the doors for research linking gut microbiota and genetic predisposition. A highly significant association with CD in two separate autophagy genes, *ATG16L1* and *IRGM* were identified and replicated during the recent whole genome association (WGA) scans. Autophagy is a recycling behavior of the cells in which they recycle surplus organelles. This process is thought to play a significant role in the normal defense against microorganisms within the cytoplasm of the cells. Therefore, autophagy serves as the mechanism to engulf and eliminate any harmful bacteria within the cell membrane. One loci identified in GWAs connecting CD with autophagy interactions has been the *ATG16L1* autophagy-related 16-like 1 (*ATG16L1*) gene on chromosome 2q37. The *IRGM* gene, is another a key contributor in autophagy. Using the Wellcome Trust Case Control Consortium (WTCCC) GWA scan a highly significant association between variation in the IRGM gene and susceptibility to CD. The genetic data for *ATG16L1* and *IRGM* clearly implicates defects in the mechanisms of autophagy with a predisposition to CD.

Recently, our International Pediatric IBD Genetics Consortium has discovered that genetic variation in *TNFRSF6B*, is associated with risk both pediatric-onset CD and UC. The *TNFRSF6B* gene product, DCR3, is a decoy receptor which is known to regulate monocyte and lymphocyte function. We have found that lymphocyte JAK/STAT signaling, and circulating cytokine levels, vary within pediatric IBD patients, after controlling for disease activity and concurrent medical therapy. These emerging data suggest that *TNFRSF6B* genotype may also regulate CD behavior and response to therapy.

Microbial Factors: Data suggest that persistent exposure to "normal resident flora" in a permeable gut might drive both initiation and relapse of clinical IBD. Moreover, the study of intestine-microbial interactions which lead to development of IBD might best be undertaken in the pediatric patient whereby influences of host fecal microbial constituents and intestinal barrier components are more easily elucidated. In addition, it is likely that effect or molecules of the innate immune system are also involved, thus, their characterization may be best achieved by studies in children with early onset, newly diagnosed disease (as described in our target cohort). Therefore, research must not fail to identify other critically important pathogens by too narrow a focus on only known pathogens or miss triggering microbes residing in colonic fecal streams in the genetically susceptible host. In addition, the characterization of gut microflora (i.e., fecal streams) and risk assessment of children with CD can optimally be done with a careful evaluation of both components – host and microbe in the index case. Successfully confirming or disproving infectious hypotheses requires that valid diagnostic tools still be developed, and that existing methods be further validated. The tools must be capable of accurately identifying speculated infectious etiologies in realistically available tissues and in fresh fecal streams that might contain the suspect microbes, and tested in well-characterized cohorts, employing well-selected comparison controls i.e., siblings and parents that have been exposed to or reside in the same environment as the index case. Systematic investigations into potential infectious etiologies or normal enteric flora triggers of IBD, particularly CD, therefore appear biologically plausible, and highly justified.

Immune markers and CD phenotypes: Existing data suggest that there are subsets of patients with differing immune responses to microbial antigens; antibodies to the E. coli outer-membrane porin C (OmpC), the Pseudomonas fluorescens CD related protein (I2), as well as Saccaromyces cerevisiae (ASCA) and autoantigens, perinuclear anti-neutrophil antibody (pANCA). Higher ASCA levels have been shown to be associated with earlier age of disease onset, both stricturing and internal penetrating disease behaviors and need for small bowel surgery. ASCA has also been shown to be associated with a more aggressive disease course among a cohort of pediatric CD patients. More recently anti-OmpC and anti-I2 have also been shown, like ASCA, to be associated with complicated disease behaviors among both adult and pediatric CD patients. None of these earlier phenotype correlation studies, however, took into account the assessment of multiple immune responses. A novel immune response, anti CBiR1 (anti-flagellin) has recently been shown to be independently associated with small bowel disease, internal penetrating and fibrostenosing disease in an adult CD cohort, but this was not demonstrated in a pediatric cohort. The results of the largest pediatric study to date demonstrated that the rate of complicated CD increases in children as the number and magnitude of immune reactivity increases. Disease progression is significantly faster in children expressing immune reactivity. Therefore, baseline immune response assessment may identify children at risk for complicating IP/FS phenotypes, in whom early, aggressive immunomodulatory therapy could be of benefit.

Granulocyte-macrophage colony stimulating factor (GM-CSF) is an endogenous cytokine required to optimal myeloid cell antimicrobial function. Recently, we have discovered that circulating levels of neutralizing GM-CSF auto-antibodies are increased in children with ileal CD, and above a threshold value are associated with a two-fold higher rate of stricturing/penetrating disease behavior and need for surgery. Consistent elevated serum GM-CSF Ab have also been associated with increased intestinal permeability, and reduced neutrophil phagocytic capacity. Therefore, we will include measurement of GM-CSF Ab as a bio-marker for complicated CD.

Intestinal Gene Expression and Early Clinical Relapse: We have recently reported that a pro-inflammatory IL-6:STAT3 dependent transcriptome is up-regulated in the colon of pediatric CD patients at diagnosis. A very similar transcriptome remains up-regulated during follow-up of patients

refractory to first line medical therapy, suggesting that a pattern of gene expression could be identified in the affected intestine at diagnosis which would predict progression to stricturing/penetrating behavior. In fact, our preliminary studies have identified a limited set of 20 genes which will identify patients likely to have early relapse, based upon colonic gene expression determined at diagnosis, prior to initiation of therapy. We will therefore determine the ability of similar studies to predict progression to stricturing/penetrating behavior, based upon the global pattern of gene expression in the ileum and colon, measured in biopsy specimens obtained at diagnosis by microarray.

D. DESIGN AND METHODS

Experimental Design

We propose a four year prospective observational cohort study design.

Subject Selection

Inclusion Criteria:

1. Males and females ≤ 16 years of age.
2. A confirmed or suspected diagnosis of CD based on standardized diagnostic criteria. The enrollment visit should occur within 30 days of diagnosis.
3. Able to provide written informed consent.
4. Have consented to have specimens tested for genetics and immune responses.
5. Access to follow up data for a minimum of 36 months after diagnosis

An upper limit of 16 years of age was selected as important inclusion criteria to maximize the number of potential subjects that would be diagnosed and followed by a pediatric gastroenterologist for at least 3 years.

Exclusion Criteria:

1. Infectious colitis

Case Ascertainment:

All patients with a confirmed or suspected diagnosis of CD are eligible for enrollment. All investigators will have a checklist form for each potential subject to confirm eligibility. A diagnosis of CD for this study will require at least 2 of the following:

1) history of abdominal pain, weight loss, short stature, malaise, rectal bleeding or diarrhea
2) characteristic endoscopic findings of discontinuous ulcerations, cobblestoning, fistula or severe perianal disease
3) radiologic features of stricture, fistula, or evidence of cobblestoning or ulceration of the mucosa
4) macroscopic appearance at laparotomy of typical bowel wall induration, mesenteric lymphadenopathy or serosal involvement showing creeping fat or other inflammatory changes
5) histopathology showing transmural inflammatory cell infiltrate or epithelial granulomas and absence of identifiable infectious agents

Patient Recruitment and Procedures

Recruitment

A total of 2000 newly diagnosed or suspected CD patients will be enrolled from the four existing consortia as well as from newly added sites (to be determined) for a total of 50 sites. The four current consortia are: Pediatric IBD Consortium (established 1999), Pediatric IBD Collaborative Research Group (established 2002), Western Regional Research Alliance for Pediatric IBD (established 2002), and Wisconsin Pediatric IBD Alliance (established 2000). They represent a total of 37 pediatric GI sites and make up the newly formed Pediatric IBD North American Collaborative Research Group (PIBD NACRG). All existing sites have the resources and capabilities to participate in all sections of this study: the collection of clinical, immune, genetic, and fecal specimens. Each participating site will obtain approval from their respective IRBs.

The study will be explained and informed consent obtained per institutional IRB requirements. While clinical data and genetic/immune markers will be obtained from all subjects, the subjects can decide to participate in a sub-study where mucosal biopsies will be obtained with an enrollment goal of 200 newly diagnosed CD patients.

Data Collection

Demographic and clinical information will be collected every 6 months or whenever the clinical data is available via uniform baseline and follow up clinical data extraction forms at all sites to ensure accuracy and consistency. The essential demographic information includes, age at diagnosis, gender, race and ethnicity, birth history, past medical history and family history. The clinical information includes characteristics of disease phenotype, surgery, growth status, and laboratory and therapeutic information. Phenotype will be defined by disease location and disease behavior. For proper classification of location, both a small bowel and large bowel examination are required. There will be 6 disease locations that patients can be categorized into and patients may be classified into more than one disease location. For proper classification of disease behavior patients must have documented radiologic, endoscopic, surgical and/or clinical evidence of disease specific behaviors as defined below. Disease behavior will be divided into 2 broad categories: non-complicating and complicating disease behaviors. Information on surgery related to development of a complication will be collected. Type of surgery (intestinal resection and or stricturoplasty), date of surgery and time to surgery after diagnosis will be documented. Tanner staging as determined by physical examination and a menstrual history where appropriate will also be recorded. Bone age when available by the method of Greulich and Pyle will be collected. Height determined on the same wall-mounted stadiometer and Tanner staging over the course of follow up, at both 12 months and 24 months after diagnosis will be recorded. Therapeutic measures used for each patient will be recorded at diagnosis and during follow-up. All data will be entered into a secured relational database which will be managed by the Data Coordinating Center (Clinipace, North Carolina).

Biologic Specimen Collection, Shipping, and Future Use of the Bio-repository

All of the bio-specimens will be bar coded and sent overnight to a central Core Laboratory at Emory University managed by Dr. Subra Kugathasan. At the time of enrollment, whole blood (7 ml for DNA/RNA, and 5 ml for serum) will be collected in all patients. Serum and DNA/RNA will be shipped overnight delivery at room temperature to the Core Laboratory at Emory. DNA/RNA will be extracted, plated and stored at the Emory DNA core facility for future genotyping studies. Serum will be aliquoted and stored for future shipping to Cedars Sinai Medical Center (CSMC) for immune response testing by Dr. Marla Dubinsky and to Cincinnati Children's Hospital Medical Center

(CCHMC) for GM-CSF Ab measurements by Dr. Lee Denson. Fresh fecal samples will be collected within 2 weeks of enrollment and will be shipped to Dr. Kugathasan's laboratory at Emory for subsequent processing within 24 hours of collection. For gene expression analysis, 6 biopsy specimens will be collected during regularly scheduled endoscopies: 3 each from the ileum and 3 each from the rectum. Four biopsies will be placed in RNALater™ and 2 biopsies will be snap-frozen then shipped to Emory. The biopsies will later be sent to CCHMC for RNA extraction, microbial, and microarray analysis by Dr. Denson.

Future use of the bio-specimens: The PIBD NACRG investigators agree entirely with the NIH policy that sharing of final data is essential for expedited translation of research results into knowledge, products and procedures to improve human health. We will follow the NIH data sharing policy as described at http://grants.nih.gov/grants/policy/data_sharing on line. The clinical data and biologic specimens collected during the study will serve as an invaluable resource for a wide spectrum of clinical and translational ancillary studies. Ancillary studies could serve to identify clinical parameters, biomarkers, immune changes or genetic patterns predictive of natural history of disease and/or specific treatment response. Currently, our planned mechanistic studies will utilize only a portion of the blood specimens that we are planning to collect. We will utilize the PIBD NACRG specimen repository at Emory University directed by Dr. Kugathasan to process, store, and catalog DNA and serum for future studies. Moreover, we will obtain immune and genetic data on each of the study subjects which may used for future studies examining genetic and immune determinants of disease phenotype. The PIBD NACRG Steering Committee will solicit, receive, and assess concept proposals for ancillary studies. All qualified investigators will be allowed access to the stored materials at the end of a pre-determined embargo period. The embargo period for each study's materials will be mutually agreed upon by the study's Steering Committee and the Pediatric Network. We anticipate that this period will be no longer than three years from accrual of the last sample or patient data.

Laboratory Methodology

Immune Markers

Serum immune responses: ASCA, ANCA, anti-CBir1, anti-OmpC and anti-I2 are determined by ELISA. Detailed methodology for the determination and characterization of the immune responses and the ELISA technique employed by the Immunobiological Institute directed by Dr. Stephan R. Targan at CSMC can be found in Appendix 1. All samples will be tested for all immune responses noted above as well as new immune responses that become of interest during the data collection period of the study. Serum GM-CSF Ab directed against glycosylated GM-CSF will also be determined by ELISA, in the laboratory of Dr. Denson at CCHMC.

Genetic Markers

All patients will be tested for mutations in *CARD15, IBD5, DLG5, MDR1, TNF, IL23R, ATG16l1, IRGM, IL6, TNFRSF6B*, and other variants using the high throughput sequenom MassARRAY iPLEX GOLD - SNP array. Additional genetic markers identified / published over the study period will also be analyzed in our cohort. This genotyping method has been used extensively in the laboratory of Dr. Kugathasan and his collaborators at Emory University for a number of studies. We will follow standard protocols for amplification and analysis.

Fecal Specimens

Fecal specimens will be stored for future testing and analysis. Proposed preliminary testing includes molecular methodologies such as gene chip and RT-PCR qualitative and quantitative analyses.

Optional sub-study:

Mucosal Biopsies
A total of 6 biopsies will be taken: 3 each from the ileum and 3 each from the rectum. Four biopsies will be placed in RNALater™ and 2 biopsies will be snap-frozen, then transported to the Core Laboratory at Emory. These specimens will be stored at appropriate temperatures and batched until transported to the Denson laboratory at CCHMC. Total RNA will be prepared to undergo microbial and microarray analysis. Patients diagnosed with CD or UC may undergo subsequent clinically indicate endoscopy procedures during the study period. In this circumstances biopsies will be obtained in exactly the same fashion as initial endoscopic procedures.

E. TOTAL NUMBER OF HUMAN RESEARCH PARTICIPANTS PROPOSED FOR THIS STUDY AT THIS SITE AND GLOBALLY.

We have based our sample size estimate upon data from a recently published cohort study involving 796 CD patients enrolled via IBD registries directed by Drs. Dubinsky, Hyams (Jeffrey – Connecticut Children's Medical Center), and Kugathasan. This study enrolled CD patients with variable durations of disease (not incident disease) and determined potential clinical (disease location), demographic (age, gender, race), genetic (*CARD15* genotype), and immune (ASCA, OmpC, CBir1, ANCA serology) predictors of stricturing/penetrating behavior. For our analysis, we considered only the group of 556 CD patients for whom immune serology was determined before the development of stricturing/penetrating behavior.

We used a logistic regression to build a probability model of developing stricturing/penetrating behavior and used parametric bootstrap to resample in a simulation study for sample size calculation. Specifically we assumed a logistic regression model

$$\log\left(\frac{p(x,t)}{1-p(x,t)}\right) = x^T\beta + \log(t),$$

where $p(x,t)$ denotes the probability of developing stricturing/penetrating behavior given some value for the predictor variables as listed above (x) and some disease duration (t). We estimated β using the group of 556 CD patients. Then, the probability of developing stricturing/penetrating behavior by $t = 36$ (36 months) for a new patient with some predictor variables (x) was given by

$$p(x,t=36) = \frac{1}{1+\exp\{-x^T\hat{\beta} - \log(t=36)\}},$$

where $\hat{\beta}$ denotes the estimate. Subsequently the observed behavior (progression to stricturing/penetrating behavior by 36 months or no progression) was decided by sampling from a Bernoulli variable with $p(x,t=36)$ for the probability of the progression.

We resampled n=1000 predictor variable (x) values from the published cohort study by Drs. Dubinsky, Hyams, and Kugathasan, and used the above probability model and sampling from Bernoulli distributions to decide observed behaviors for the 1000 resample observations. We split them into an index and a validation set of 500 each and used logistic regression model to build a prediction model based on the index set and to validate the prediction model using the validation set.

Based on 1000 simulations, we estimated the following predictive characteristics of the model:

	c	Gamma	Tau-a	Positive Predictive Value (PPV) = 1 – Rate of False Positives	Negative Predictive Value (NPV) = 1 – Rate of False Negatives
Mean [SD] (90% CI)	0.713 [0.030] (0.67, 0.80)	0.428 [0.061] (0.34, 0.61)	0.0923 [0.014] (0.07, 0.13)	0.216 [0.051] (0.134, 0.303)	0.865 [0.015] (0.865, 0.914)

This table suggests that if the prospective data to be collected for the proposed study is not very different from the published cohort study by Drs. Dubinsky, Hyams, and Kugathasan, we anticipate to build a prediction model whose accuracy and predictive power in terms of c-statistics and PPV/NPV respectively can be estimated with error < 0.1, conservatively speaking, with the proposed plan of accruing a total of n=1000 patients and splitting them into equal size index and validation (or test) sets of 500 each.

For the sub-study involving mucosal biopsy collection, we will enroll 200 CD patients who are undergoing their initial diagnostic evaluation and have inflammatory disease behavior. We anticipate that 20 percent of these, or 40, will progress to stricturing/penetrating behavior during follow-up. We will determine the global pattern of gene expression in the intestinal biopsies from these patients, and will use these data to identify a minimal gene list (n=20) which will distinguish CD patients with inflammatory disease behavior during follow-up, from those who progress to stricturing/penetrating behavior. Using class prediction analysis algorithms in GeneSpring™, this will provide 20 subjects with inflammatory behavior, and 20 subjects who progress to stricturing/penetrating behavior, for the index set, and the same number for validation set.

F. DRUGS OR PROCEDURES

There are no drugs involved in this study. Procedures include prospective data collection from medical records (every 6 months), venipuncture, stool collection, and obtaining 6 additional gastrointestinal mucosal biopsies for research purposes during a routine endoscopic procedure as dictated by clinical indication. The frequency of events is listed in the events chart below. Subjects will be reimbursed $25 per study visit.

VISIT	**1**	**2**	**3**	**4**
Activity	Screening	12 Months	24 Months	36 Months
Informed consent process	X			
Venipuncture (12 ml screening, 12 ml annually)	X[A]	X[B]	X[B]	X[B]
Stool Collection	X[C]			
Mucosal Biopsy Collection (sub-study)[E]	X[D]			

A. 12 ml of blood (7 ml for DNA/RNA and 5 ml for serum) will be drawn at the screening visit for genetic and immune testing

B. 12 ml of blood will be drawn annually for immune testing (7 ml for DNA/RNA and 5 ml for serum)

C. 1 stool sample will be collected at screening or at follow-up if patient unable to provide at screening

D. 6 biopsy samples will be taken during regularly scheduled procedure, at time of diagnosis / screening
E. Additional biopsies will be obtained if repeat endoscopy is clinically indicated.

G. RISK CATEGORY:

The risk category for subjects who enroll into the study will be categorized as minimal risk. The sub-study involving mucosal biopsy collection will also fall into the minimal risk category based on Federal Regulation 45 CFR 46.404**.** Although some institutions may regard obtaining mucosal biopsies as more than minimal risk, Emory and several other institutions have recently approved that obtaining mucosal biopsies for research during the routine endoscopic procedures are in the minimal risk category.

H. RISKS AND THE PRECAUTIONS WHICH WILL BE TAKEN TO MINIMIZE RISK EXPOSURE

The risks of this study are limited to those inherent with mucosal biopsy collection, venipuncture, and breach of confidentiality. Venipuncture may cause some pain, bleeding or bruising at the spot where the needle enters the body. Rarely, taking blood may cause fainting or infection. The primary risk associated with a mucosal biopsy is bleeding from the area of biopsy. Significant bleeding from the biopsy site is significantly uncommon, but possible (1 in 1,000). Even more rarely (1 in 3,000), a perforation can be made at the site of biopsy which would require surgery to repair. To minimize these risks, standard practice is to not take a biopsy of large vessel, not more than one biopsy from a given mucosal site and not to over distend the bowel lumen while taking biopsies. Further more, mucosal biopsies are not painful as the bowel lining does not have pain fibers to sense pain. In addition, all the above mentioned risks apply regardless if the biopsy is done for donating tissue to this study or for diagnostic purposes. Venipuncture and mucosal biopsy sampling will be performed by medical and nursing staff with expertise in performing these procedures in children.

I. PROVISION FOR THE PROTECTION OF PRIVACY OF SUBJECTS AND TO MAINTAIN THE CONFIDENTIALITY OF DATA

Risk of breach of confidentiality will be minimized as only site specific PIs and their research staff will have access to patient identifiers. All data will be sent to the Data coordinating center (Clinipace) to a secure data management platform ('TEMPO'), the software used by Clinipace. Tempo is a highly secure web-based system which generates a unique identifier as individual records are created. No PHI is included in the identifier. All data transfers from sites are encrypted using standard 128-bit encryption and the data itself is protected by Checkpoint's Firewall 1-UTM. Clinipace does not own rights to the data, nor claims any rights to study, patient, or clinician data. CCFA and clinicians participating in this CCFA supported project retain ownership rights to all data stored in Clinipace.

Study samples will be stored in locked refrigerators/ freezers at the Core Laboratory at Emory and will only be linked to phenotype by study code. All potential identifiers will be removed from the samples by the site specific investigators/staff prior to shipping to the Core Laboratory.

HIPAA AUTHORIZATION FOR USE/DISCLOSURE OF PROTECTED HEALTH INFORMATION FOR RESEARCH: We will obtain written authorization (permission) before we may use or disclose (release) the subject's "protected health information" (sometimes referred to as "PHI") related to the study. This form provides that authorization and ensures that the subject and subject's parents are properly informed of how this information will be used or disclosed.

J. PROVISIONS FOR MONITORING DATA TO ENSURE THE SAFETY OF SUBJECTS AND ADDITIONAL SAFEGUARDS TO PROTECT THE RIGHTS AND WELFARE OF SUBJECTS WHO ARE LIKELY TO BE VULNERABLE

This study is a non-intervention study. Therefore, no formal data safety monitoring board will be appointed. **Our data and safety monitoring plan will be as follows:** The risks associated with this study are associated with two procedures: venipuncture and biopsy during colonoscopy. Adverse events associated with venipuncture will be monitored for by the site study coordinator, and reported to the site PI. Colonoscopies will be performed for clinical diagnostic indications in patients with suspected CD. Adverse events will be monitored for, and managed by, the primary gastroenterologist performing the colonoscopy. These adverse events will be reported to the site PI, who will then report these to the local IRB within the time frame stipulated by the local IRB regulations. Our studies will enroll a vulnerable population, pediatric subjects. Please see section N. below for the procedures we will follow during the consenting process to protect their rights and welfare. Venipuncture and mucosal biopsy sampling will be performed by medical and nursing staff with expertise in performing these procedures in children. We have used these procedures in prior and current similar studies, and anticipate that they will be quite effective in protecting the rights and welfare of the pediatric subjects.

K. ANTICIPATED BENEFITS ASSOCIATED WITH THE PROTOCOL TO HUMAN RESEARCH PARTICIPANTS AND SOCIETY

We expect serotyping, genotyping and microbial factors will identify about 20% of children with CD at risk of developing complicated disease at 36 months follow up (out of proposed 1100 children, 200 will be at risk). Further analysis of genetic polymorphisms and immune markers with modeling will identify about half of these children (n=165) who may benefit with aggressive medical therapy based on our modeling (Aim 2). This new knowledge is likely to benefit future CD patients.

L. STOPPING POINTS THAT WOULD NOT ALLOW THE STUDY TO CONTINUE AS PROPOSED (point in time when it has been determined that the objectives have been met; it has been determined that the objectives cannot be met or the accumulated data indicates that the risks exceed the benefits of the study)

There are no formal stopping rules apply to this proposal since it is an observational and validation study without any intervention in a prospective cohort.

M. IS THERE A DATA SAFETY MONITORING BOARD IN PLACE? WHO ARE IT'S MEMBERS? HOW OFTEN DO THEY MEET?

There is no DSMB for this study. There is a Data and Safety Monitoring Plan (see section J above).

N. CONSENTING PROCESS

For those patients meeting inclusion criteria, the study purpose, procedures, costs, risks, benefits and alternatives to participation will be thoroughly explained and presented to the patient and their family by the site specific designated study coordinator and/or investigator, and subjects will be screened for willingness to participate. Once patients and families have had enough time to consider participation and have expressed a willingness to participate, informed consent will be obtained from the parents or

legal guardian and child assent when appropriate based on age and institutional IRB requirements.

O. PROCEDURES TO BE EMPLOYED IN ANALYZING DATA AND THE ANTICIPATED SIGNIFICANCE OF THE PROPOSED STUDY

Analysis: The primary endpoint for this study will be the development of stricturing/penetrating behavior within 36 months from enrollment, in CD patients who at enrollment have inflammatory behavior. The secondary endpoint will be the need for surgery in these patients. Based upon our preliminary data, we anticipate that out of an inception cohort of 1100 CD patients with inflammatory behavior at diagnosis, 200 will progress to stricturing/penetrating behavior. We **hypothesize** that the following parameters, measured at diagnosis, will predict the stricturing/penetrating outcome: age at diagnosis, gender, intestinal disease location, *CARD15* genotype, ASCA serology, OmpC serology, CBir1 serology, ANCA serology. We will also determine *ATG16L1*, *IRGM*, and *TNFRSF6B* genotype, and GM-CSF Ab serology, as exploratory variables.

We will split our cohort (n=1000) into equal size index and validation (or test) sets of 500 each. We will split the cohort as they are enrolled using a randomized block design with blocks of 2, provided by the DCC, Clinipace. This will ensure equal numbers of approximately equal duration of follow-up between the index and validation cohorts. We will use the index set and logistic regression to build a prediction model for the probability of developing stricturing/penetrating behavior and use the validation set to evaluate the prediction model. We will use the c statistic and associated SD to summarize the operating characteristics of the prediction model on the index case set, and positive predictive value (PPV), negative predictive value (NPV) and their SD to summarize the operating characteristics of the prediction model on the validation set. Our preliminary analysis suggests that we build a prediction model whose predictive power in terms of c-statistics can be estimated with error < 0.1.

A secondary endpoint for this study will be the time until the development of stricturing/penetrating behavior from enrollment. We will use Kaplan-Meier estimator and log-rank test to estimate and test the probability of no progression to complication (stricturing and/or penetrating disease behavior) by each of the independent variables. To estimate and test the effect of each of the independent variables conditioning on the rest of the variables, we will use Cox proportional hazards regression.

Descriptive statistics will be used to define the relative frequency of each of the independent and dependent variables, for the study population as a whole, and for each of the study sites. We will also determine the relative frequency of corticosteroid, immunomodulator, and biologic therapy for the over all cohort, and for each study site. Differences for these between the sites will be determined by ANOVA for continuous variables, and χ^2 test for dichotomous variables with Bonferroni correction for multiple comparisons. The Kruskal-Wallis test for continuous variables will be used when the normality of the data is suspected. These data will determine whether these parameters which may affect stricturing/penetrating behavior differ between study sites, or between the index and validation cohorts.

For the sub-study involving mucosal biopsy collection, we will enroll 200 IBD patients who are undergoing their initial diagnostic evaluation and have inflammatory disease behavior. As clinically indicated, biopsies are taken for diagnostic purposes in each subject, and we will use the same biopsy forceps with same technique for obtaining biopsy samples for the research purposes. A total of 6 biopsies will be taken: 3 each from the ileum and 3 each from the rectum. We anticipate that 20 percent of these, or 40, will progress to stricturing/penetrating behavior during follow-up. Microarray analysis will be performed to determine the pattern of gene expression for the CD genomic signature (CDGS)

in the 40 subjects who experience complicated disease behavior, and in 40 matched for age, gender, and disease location and severity at diagnosis. We will perform an initial class prediction analysis with the first forty CD patients (20 with progression to stricturing/penetrating behavior and 20 with inflammatory behavior), to define the minimal gene set which will distinguish stricturing/penetrating behavior (CD_{SP}) from inflammatory behavior (CD_I) with at least 95% accuracy by cross validation; this will comprise the training or index set. We will begin by determining the global pattern of gene expression for the CDGS which we have established contains predictive genes in our preliminary studies. We will make use of the training/test set application within GeneSpring™ software to perform this analysis. This is similar to the method recently used to define a 14 gene set which will distinguish between CD and UC with 94% accuracy using RNA from PBMC. We will then test the accuracy of the gene set identified during the training phase using samples obtained from the second forty CD patients in the cohort; this will comprise the test set.

Significance: Development of the proposed risk stratification model will provide for timely identification of pediatric CD patients likely to have progressive, stricturing/penetrating, disease behavior. In the short run, this will form the basis for a randomized controlled trial of early biologic therapy in high risk patients. In the long run, this will provide a novel clinical tool for providing therapy more targeted to the individual patient's likelihood of more severe disease.

P. FINANCIAL RELATIONSHIPS

This study is being funded by Crohn's and Colitis Foundation of America, a non-profit organization. The study centers are receiving funds to cover a part of the costs of conducting the study. Investigators have no equity/stock interests in any organization associated with this study.

Q. BIBLIOGRAPHY (list pertinent literature references)

Langholz E, Munkholm P, Krasilnikoff PA, Binder V. Inflammatory bowel diseases with onset in childhood. Clinical features, morbidity, and mortality in a regional cohort. Scand J Gastroenterol 1997;32:139-47

Gupta N, Cohen SA, Bostrom AG, Kirschner BS, Baldassano RN, Winter HS, Ferry GD, Smith T, Abramson O, Gold BD, Heyman MB. Risk factors for initial surgery in pediatric patients with Crohn's disease. Gastroenterology 2006;130:1069-77

Griffiths AM, Nguyen P, Smith C, MacMillan JH, Sherman PM. Growth and clinical course of children with Crohn's disease. Gut 1993;34:939-43

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599-603

Farrell RJ, LaMont JT. Microbial factors in inflammatory bowel disease. Gastroenterol Clin North Am 2002;31:41-62

Hyams J, Crandall W, Kugathasan S, Griffiths A, Olson A, Johanns J, Liu G, Travers S, Heuschkel R, Markowitz J, Cohen S, Winter H, Veereman-Wauters G, Ferry G, Baldassano R. Induction and maintenance infliximab therapy for the treatment of moderate-to-severe Crohn's disease in children. Gastroenterology 2007;132:863-73; quiz 1165-6

Cosnes J, Cattan S, Blain A, Beaugerie L, Carbonnel F, Parc R, Gendre JP. Long-term evolution of disease behavior of Crohn's disease. Inflamm Bowel Dis 2002;8:244-50

Ahmad T, Armuzzi A, Bunce M, Mulcahy-Hawes K, Marshall SE, Orchard TR, Crawshaw J, Large O, de Silva A, Cook JT, Barnardo M, Cullen S, Welsh KI, Jewell DP. The molecular classification of the clinical manifestations of Crohn's disease. Gastroenterology 2002;122:854-66.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Diversity of the human intestinal microbial flora. Science 2005;308:1635-8

Landers CJ, Cohavy O, Misra R, Yang H, Lin YC, Braun J, Targan SR. Selected loss of tolerance evidenced by Crohn's disease-associated immune responses to auto- and microbial antigens. Gastroenterology 2002;123:689-99

Dubinsky MC, Lin YC, Dutridge D, Picornell Y, Landers CJ, Farrior S, Wrobel I, Quiros A, Vasiliauskas EA, Grill B, Israel D, Bahar R, Christie D, Wahbeh G, Silber G, Dallazadeh S, Shah P, Thomas D, Kelts D, Hershberg RM, Elson CO, Targan SR, Taylor KD, Rotter JI, Yang H. Serum Immune Responses Predict Rapid Disease Progression among Children with Crohn's Disease: Immune Responses Predict Disease Progression. Am J Gastroenterol 2006;101:360-7

Targan SR, Landers CJ, Yang H, Lodes MJ, Cong Y, Papadakis KA, Vasiliauskas E, Elson CO, Hershberg RM. Antibodies to CBir1 flagellin define a unique response that is associated independently with complicated Crohn's disease. Gastroenterology 2005;128:2020-8.

Appendix 1

IMMUNE RESPONSE TESTING

ASCA/ANCA ELISA

ASCA: The samples will be analyzed by ELISA with phosphopeptidomannan extracted from Saccharomyces cerevisiae serving as the antigen. Briefly, plasma diluted 1:80 (for IgA detection) or 1:800 (for IgG detection) will be added to plates previously coated with mannan at 100ug/ml. After incubation and washing, alkaline phosphatase labeled goat anti-human IgA and IgG are added to their respective plates. Finally, after another incubation and wash, substrate (p-nitrophenyl phosphate) is added and color change is detected at 405nm. All samples are compared to standard positive control samples and expressed as ELISA units (EU).

ANCA: The samples will be quantitatively analyzed by ELISA and ELISA positive samples will be further characterized qualitatively by immunofluorescence binding pattern including DNase treatment of samples with pANCA binding. For ELISA analysis, microtiter plates are coated with a monolayer of neutrophils (25,000/well) and air-dried, fixed with 100% methanol, dried again and stored at -20oC. For use, the plates are blocked for non specific binding by 0.5% bovine serum albumin in phosphate buffered saline (BSA/PBS), the blocking material is discarded and samples at a 1:100 dilution in BSA/PBS are added. After incubation and washing, alkaline phosphatase labeled goat anti-human IgG (gamma chain specific) is added. Finally, after another incubation and wash, substrate (pnitrophenyl phosphate) is added and color change is detected at 405nm. All samples are compared to standard positive control samples and expressed as ELISA units (EU). For indirect immunofluorescence analysis of ANCA, slides are prepared by cytocentrifugation of 100,000 neutrophils. The slides are air-dried and fixed in 100% methanol, then air-dried and stored at -20oC. For use, the slides are rehydrated in PBS and samples diluted 1:20 in BSA/PBS are added. After washing, fluoroscein labeled goat F (ab)'2 anti human IgG (gamma chain specific) is added. The slides are washed again and evaluated by fluorescence microscopy. For evaluation of DNase sensitivity of pANCA patterns, slides are pretreated with 100U/ml of RNase free DNase for 30 minutes. Slides are stained as above with a pair of DNase-treated and untreated slides used for each sample.

CBir1 ELISA: ELISA analysis of anti-CBir1 is performed as previously described, but using NH2-terminal fragment of CBir1 (147aa) without knowledge of diagnosis or other serology results. Briefly, ELISA plates are coated overnight with 100ng/well of CBir1, then blocked with 1% BSA in PBS for 2 hours. Plates are washed and serum added at a 1:200 dilution in 1% BSA-PBS for a 30 minute incubation. After washing, horseradish peroxidase conjugated anti-human IgG at a 1:10,000 dilution is added and incubated for 30 minutes. After another wash, the plates are incubated with tetramethylbenzidine substrate for 15 minutes. The reaction is stopped with 1 N sulfuric acid and read at 450nm. Positive is defined as the mean + 2 SD of the healthy controls. For Cohort 2 and the longitudinal cohorts and phenotype cohorts, this assay is modified to be more similar to the ANCA, OmpC and I2 protocols: alkaline phosphatase is substituted as the secondary conjugate and incubated for 1 hour followed by paranitrophenyl phosphate as substrate for 30 minutes.

OmpC/I2 Purification: Trimeric **OmpC** is biochemically purified from an $OmpF^{-/-}/OmpA^{-/-}$ disruptive insertion mutant E. coli K12 (provided by R. Misra). Mutant E. coli glycerol stocks are inoculated into 10-20ml of Luria Bertani broth supplemented with 100ug/ml Streptomycin (LB-Strep), and culture vigorously at 37°C for ~8 hours to log phase followed by expansion to 1 liter in LB-Strep over 15 hours at 25□C. Cells are 50 harvested by centrifugation, washed twice with 100 ml of ice cold

20mM Tris-Cl pH 7.5, and resuspend in cold spheroplast forming buffer (20 mM Tris-Cl pH 7.5, 20% Sucrose, 0.1M EDTA pH 8.0, 1mg/ml Lysozyme). Spheroplasts are allowed to form for 1 hour on ice with occasional mixing, and then lysed by 14 fold dilution into ice cold 10mM Tris-Cl pH 7.5, 1mg/ml DNase-I, and vigorous vortexing followed by pulse sonication (4x30seconds, On time=1 second at high power). Cell debris is by low speed centrifugation, and membrane preparation collected by ultra centrifugation at 100,000g in a swing bucket rotor. Membrane pellet is resuspended by homogenizing into 20mM Tris-Cl pH 7.5, and extracted for 1 hour in 20mM Tris-Cl pH 7.5 + 1% SDS by rotating at 37°C. Pre-extracted membrane preparation is then pelleted by ultracentrifugation and resuspended by homogenizing into 20mM Tris-Cl pH 7.5 as above, and OmpC is extracted for 1 hour rotating at 37°C with 20mM Tris-Cl pH 7.5, 3%SDS, and 0.5M NaCl. Membrane is then pelleted by ultracentrifugation and the supernatant containing trimeric OmpC is collected. SDS is removed from OmpC preparations by detergent exchange dialysis against >10,000 volumes of 0.2% triton x100 followed by dialysis against >10,000 volumes Tris-Cl pH 7.5. Purified OmpC is quantified using the Bradford reagent (Biorad, Hercules, CA) and purity of >95% is validated by SDS-PAGE and Silver staining (Biorad). Purified protein is aliquoted and stored at –20ºC until used.

The 100 amino acid open reading frame (ORF) of **I2** is subcloned into pGEX-KG and expressed in E. coli XL-1 blue (Stratagene, La Jolla, CA). I2-GST fusion protein is present as an inclusion body, and purified according to manufacturer's instructions by differential solubilization in 0.1% sodium dodecyl sulfate (SDS). The glutathione-S-transferase (GST) control is produced with unmodified pGEX-KG and XL-1 blue cells, and is present about 50% in the soluble and inclusion body fractions. The latter is purified exactly as I2-GST, and the former is purified by G-Sepharose affinity chromatography. All protein preparations are >90% pure by SDS polyacrylamide gel electrophoresis (PAGE) and Coomassie blue protein staining.

DETERMINATION AND CHARACTERIZATION OF THE OmpC/I2 RESPONSE

Human IgA antibodies that bind I2 or OmpC will be detected by direct ELISA assays. Plates (Greiner, USA Scientific, Ocala, FL) will be coated overnight at 4 C with 100 µl/well of GST alone and I2-GST (5µg/ml) or OmpC (0.25µg/ml) in borate buffered saline, pH 8.5. After three washes in 0.05% Tween 20 in phosphate buffered saline (PBS), the plates will be blocked with 150 µl/well of 0.5% bovine serum albumin in PBS, pH 7.4 (BSA-PBS) for 30 minutes at room temperature (RT). The blocking solution will then be discarded and 100 µl/well of sera diluted 1:100 will be added and incubated for 2 hours at RT. The plates will be washed as before and alkaline phosphatase conjugated goat antihuman IgA (∟-chain specific, Jackson ImmunoResearch, West Grove, PA) at a dilution of 1:1000 in BSA-PBS will be added for 2 hours at RT. The plates will be washed three times with 0.05% Tween 20 in phosphate buffered saline followed by another three washes with Tris buffered normal saline, pH7.5. Substrate solution (1.5mg/ml disodium P-nitrophenol phosphate (Amresco, Solon, OH), 2.5mM MgCl2, 0.01M Tris, pH 8.6) will be added at 100 ∟l/well and color will be allowed to develop for one hour at which time the plates will be read at 405nm. Nonspecific binding of sera to GST alone (typically <0.1) will be subtracted from raw values of I2-GST binding to obtain I2 specific absorbances. Levels will be determined relative to a standard consisting of serum obtained from a well-characterized CD patient. Results will be expressed as ELISA units (EU/ml). Sera with antibody levels exceeding the normal reference range value will be termed positive.