Missense Mutation in *SCGN* (Secretagogin) as a Possible Cause of Ulcerative Colitis in Three Siblings

Ernesto M. Llano[1], Lillienne Chan[2], Steven M. Harrison[3], Haiying Li[1], Luis Sifuentes-Dominguez[2], Jason Y. Park[4], Jonathan C. Cohen[1], Linda A. Baker[3], Ezra Burstein[1,5]

Departments of Internal Medicine[1], Pediatrics[2], Urology[3], Pathology[4], and Molecular Biology[5], UT Southwestern Medical Center, Dallas, Texas

Background

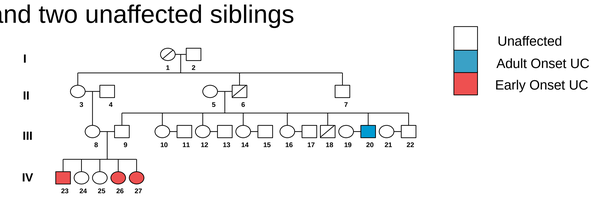
- Inflammatory bowel disease (IBD) consists of two disorders, Crohn's disease and ulcerative colitis (UC)
- IBD affects 1.4 million Americans and has a peak incidence between 15 and 30 years of age
- Genetic factors contribute to the development of these disorders, but all culprit genes have not been identified

Purpose

To further our understanding of genes responsible for IBD heritable risk, we examined a consanguinous family with high incidence of childhood-onset UC

Pedigree under study

- Consanguinous family of Mexican origin
- Parents of affected children are first cousins
- Three kids affected with childhood-onset severe UC and two unaffected siblings



Ranges of interest			Genes Contained (total 64)	Incomplete coverage in Exome sequence (total 12)
Chr	Start	End		
6	15,047,472	15,570,006	JARID2 (DTNBP1) (DTNBP1) MYLIP GMPR ATXN1	One exon (16,238,872-16,239,068)
6	15,571,356	16,884,088	STMN2/1 RBMZ4 CAP2 FAM8A1 NUP153 KIF13A NHLRC1 TPMT (KDM1B)	All five exons (17,102,443-17,131,163) One exon (17,652,063-17,652,160) One exon (17,987,600-17,987,759) One exon (18,122,514-18,122,908)
6	18,262,607	19,682,599	(DEK) RNF144B AK098665 AK097585	
6	19,686,123	19,803,768	None	
6	19,804,188	22,312,287	(AK092541) ID4 MBOAT1 E2F3 CDKAL1 SOX4 PRL	One exon (19,837,965-19,838,161) One exon (20,403,950-20,404,146) One exon (21,595,339-21,596,047)
6	23,584,375	26,148,311	NRSN1 DCDC2 KAAG1 MRS2 GPLD1 ALDH5A1 KIAA0319 TDP2 ACOT13 C6orf62 GMNN FAM83B CMAHP BC029534 LRRC16A SCGN HIST1H2AA HIST1H2BA DQ588578 HIST1H2AOS1 SLC17A4 SLC17A1 SLC17A3 SLC17A2 TRIM38 HIST1H1A HIST1H1A HIST1H4A HIST1H4B HIST1H3B HIST1H2AB HIST1H2BB HIST1H3C HIST1H1C HFE HIST1H4C HIST1H3T HIST1H3BC HIST1H2AC	One exon (24,357,791-24,358,185) One exon (24,512,029-24,512,126) One exon (24,564,401-24,564,597) One exon (24,839,775-24,839,872)

Ranges of interest			Genes Contained (total 76)	Incomplete coverage in Exome sequence (total 9)
Chr	Start	End		
12	71,016,157	72,070,710	(PTPRB) PTPRR TSPAN8 LGR5 2PC3H1 (THAP2) (TBC1D15) TPH2 TRHDE	One exon (71,182,533-71,182,686) One exon (72,037,835-72,038,143)
12	72,320,431	73,669,855	LOC100507377 ATXN7L3B KCNC2 AK093193 CAPS2 GLIPR1L1	
12	73,671,276	75,891,939	GLIPR1L2 (GLIPR1) (KRR1) (CSRP2) E2F7 NAV3 BC045810 BC047615 (SYT1) PAWR PPP1R12A OTOGL PTPRQ MYF6 MYF5 BC037383 AK022997 LIN7A ACSS3 (PPFIA2)	Two exons (75,796,784-75,796,937; 75,824,407-75,825,024) All 56 exons (80,750,134-80,772,870) One exon (81,991,649-81,991,957)
12	77,249,400	79,204,389		
12	79,302,419	82,138,159		
12	82,147,487	83,157,896	PTPRB CCDC59 METTL25 (TMTC2)	
12	83,541,966	88,441,286	SLC6A15 BC045599 TSPAN19 LRRIQ1 ALX1 RASSF9 NTS MGAT4C C12orf60 C12orf29 KITLG LOC728084 DUSP6 (POC1B) (POC1B) ATP2B1 CCER1 EPYC KERA LUM DCN	One exon (91,497,940-91,498,093)
12	88,594,159	89,910,070		
12	89,915,484	91,765,486		
12	91,766,720	93,152,115	LOC256021 BTG1 BC044741 CLUL1/DS	Two exons (92,816,287-92,816,440; 92,815,165-92,815,782) One exon (92,821,821-92,821,974)
12	93,251,750	95,410,845	CLUL1 C12orf74 (PLEKHG7) (EEA1) NUDT4 UBE2N MRPL42 SOCS2 ORAOO AK13091 PLXNC1 CCDC41 TMCC3 NDUFA12	Two exons (92,821,821-92,821,974; 93,792,511-93,792,664)

Figure 2. All genes found in the intervals of homozygosity shared only by the three probands. Some genes found in these regions of interest were incompletely covered by exome sequencing and are being further examined with whole genome sequencing.

Methods

- Exome sequencing of all children (affected and unaffected) was performed at the McDermott Center next generation DNA sequencing core, using Illumina HiSeq 2500 sequencer and the TrueSeq exome capture library.
- Given the consaguinity in this kindred, areas of homozygosity were examined using a SNP array (Infinium HumanCoreExome Beadchip SNP array). This standard Illumina SNP chip types 264,909 tag SNP markers (244,593 in exons).
- Potential variants of interest were confirmed by PCR amplification / Sanger sequencing and population frequencies for variants of interest were directly determined using allelic discrimation assays in DNA samples of Hispanic individuals in the Dallas Heart Study.

Results

- Two areas of homozygosity, in Chr6 and Chr12, were shared by the 3 affected probands and not by their unaffected siblings, encompassing a total 31.9 Mb of shared sequence (Figure 1). Given the inheritance pattern in the pedigree, we speculated that a homozygous recessive mutation in a gene contained within these intervals should be responsible for the phenotype. Thus, a total of 140 potential genes were implicated (Figure 2).
- Exome sequencing results were prioritized using the following criteria: 1) The variants had to affect the coding regions of both alleles in the 3 probands and not their 2 unaffected siblings, 2) the variants had to be rare in the general population, using the 1000 genome (oneKG) database as a reference. Afterwards, we cross-referenced the identified potential culprit variants found in the exome sequence analysis against the unique areas of homozygosity shared by the probands (Figure 3). The top candidate change was in *SCGN*, which had a coding variant (c.433G>A/p.R77H) not previously found in the oneKG database or in the dbSNP database. Sanger sequencing confirmed that both parents were carriers, the three probands were homozygous, and one sibling was a carrier while the other was wild-type.
- Given the paucity of Hispanic reference genomes, we directly genotyped 2000 individuals of Hispanic descent for this variant, and found only one heterozygote carrier. This frequency would predict that only 1:16 million Hispanics is homozygous for this variant (1/2000 x 1/2000 x 1/4).
- A second coding variant in KIF13A was less likely implicated since 4% of Caucasians are homozygous for this change.
- *SCGN* encodes a 276 amino acid, calcium-binding protein. The protein is expressed in tissues of neuroendocrine lineage, such as pancreatic β cells and intestinal enteroendocrine cells.
- The coding change found in these patients is located in one of the calcium binding domains (EF hand 2). Immunohistochemical analyses of the patients' colon biopsies did not demonstrate a decrease of the presence of SCGN when compared to controls, suggesting that the mutation does not affect expression, but may have a functional effect yet to be determined.

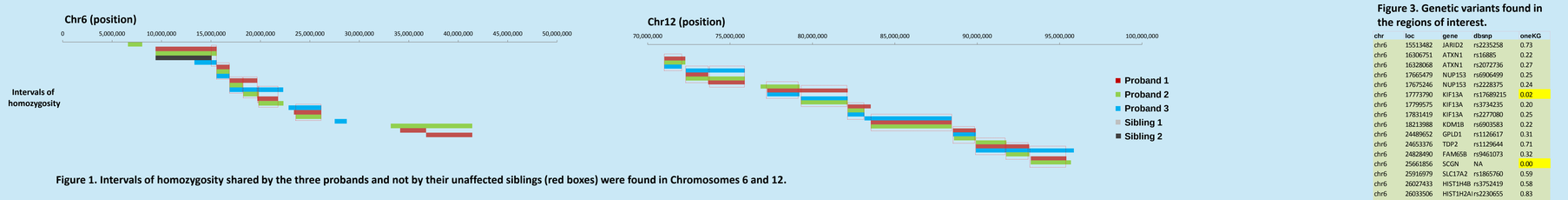


Figure 1. Intervals of homozygosity shared by the three probands and not by their unaffected siblings (red boxes) were found in Chromosomes 6 and 12.

Conclusions and Future Directions

- A very rare variant of *SCGN* was found to be the most likely culprit of UC in these patients
- Given various gaps in the exome sequencing results affecting a total of 21 genes of interest, confirmatory whole genome sequencing is currently in progress
- We are building a molecular and cellular model of the R77H mutation to identify if this coding change alters the function of the protein
- If a functional impairment of SCGN is confirmed, it could represent the first description of enteroendocrine dysfunction playing a role in IBD pathophysiology

Figure 3. Genetic variants found in the regions of interest.

chr	loc	gene	dbsnp	oneKG
chr6	15513482	JARID2	rs2235258	0.73
chr6	16306751	ATXN1	rs16885	0.22
chr6	16328068	ATXN1	rs2072736	0.27
chr6	17665479	NUP153	rs6906499	0.25
chr6	17675246	NUP153	rs2228375	0.24
chr6	17773790	KIF13A	rs176892151	0.02
chr6	17799575	KIF13A	rs3734235	0.20
chr6	17831419	KIF13A	rs2277080	0.25
chr6	18213988	KDM1B	rs6903583	0.22
chr6	24489652	GPLD1	rs1126617	0.31
chr6	24653376	TDP2	rs1129644	0.71
chr6	24828490	FAM65B	rs9461073	0.32
chr6	25661856	SCGN	NA	0.00
chr6	25916979	SLC17A2	rs1865760	0.59
chr6	26027433	HIST1H4B	rs3752419	0.58
chr6	26033506	HIST1H2AI	rs2230655	0.83
chr6	26045929	HIST1H3C	rs3752416	0.80
chr6	26056549	HIST1H1C	rs10425	0.80
chr6	26091179	HFE	rs1799945	0.13
chr12	71029733	PTPRB	rs2584021	0.34
chr12	71526593	TSPAN8	rs2270587	0.23
chr12	72372862	TPH2	rs7305115	0.52
chr12	72416235	TPH2	rs4290270	0.55
chr12	77438436	E2F7	rs2242384	0.25
chr12	78400884	NAV3	rs34276383	0.17
chr12	78530979	NAV3	rs1852464	0.59
chr12	79611374	SYT1	rs2037743	0.67
chr12	80899901	PTPRQ	rs11114486	0.60
chr12	80900397	PTPRQ	rs12824064	0.61
chr12	81627238	ACSS3	rs1921038	0.22
chr12	89744773	DUSP6	rs7700887	0.28
chr12	89916811	GALNT4,P	rs2230283	0.31
chr12	89917518	GALNT4,P	rs2230281	0.68
chr12	90028901	ATP2B1	rs1050395	0.17
chr12	92818786	CLUL1	rs1515565	0.66
chr12	93147907	PLEKHG7	rs9669519	0.82
chr12	93150102	PLEKHG7	rs924326	0.81
chr12	94543506	PLXNC1	rs2230754	0.33
chr12	94645255	PLXNC1	rs2230757	0.71
chr12	94972290	TMCC3	rs2270893	0.52
chr12	94975799	TMCC3	rs3747553	0.57
chr12	94976084	TMCC3	rs3747552	0.58
chr12	94976159	TMCC3	rs3747551	0.34