

Alternative Splicing Signature in human Crohn's Disease

Daowei Li

The People's Hospital of Liaoning Province

Yue Tan (✉ tanyue2006@163.com)

Shengjing Hospital of China Medical University <https://orcid.org/0000-0002-0373-2202>

Primary research

Keywords: Alternative Splicing, Crohn's disease, post-transcriptional regulation

Posted Date: January 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-113422/v2>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

Although hundreds of risk loci for Crohn's disease (CD) have been identified, the underlying pathogenesis of CD remains unclear. Recently, evidence has shown that aberrant gene expression in CD patients' colon tissues was associated with progression of CD. We reasoned that post-transcriptional regulation, especially alternative splicing (AS), may also play important roles in the pathogenesis of CD. Herein we re-analyzed a public mRNA-seq data from NCBI GEO dataset (GSE66207), and identified approximately 3,000 unique AS events in CD patients compared to the healthy controls. "Lysine degradation" and "Sphingolipid metabolism" are the top two AS events enriched terms in CD patients. In a validation study, we also sequenced 8 subjects and demonstrated that key genes which were previously linked to CDs, such as IRF1, STAT3, also had significant AS events in CDs. In conclusion, our study provided a landscape of AS events in CD, especially as the first study focused on a Chinese cohort. Our data suggested that dysregulation of AS may be a new mechanism that contribute to the pathogenesis of CD.

Introduction

Crohn's disease (CD), as an inflammatory bowel disease (IBD), was recognized to be possibly caused by dysregulated mucosal immune responses to the gut flora, in these genetically susceptible patients (Simmons, 2010). To date, more than 200 risk loci for IBD have been identified, which are also enriched in multigenic regulatory modules (Hong et al., 2016; Momozawa et al., 2018). Although genome-wide association studies as well as subsequent meta-analysis could explain the underlying pathophysiology of CD to some extent, the pathogenesis of CD remains not fully understood. Recently, studies showed that dysregulated gene expression in the gut mucosa of CD patients may determine the initiation and progression of CD (van Beelen Granlund et al., 2013). A following study also showed there exists a big difference of gene expression in inflamed and noninflamed intestinal mucosa of CD patients when compared with the healthy mucosa of controls (Hong et al., 2017). Moreover, a study that focused on integrated analysis of micro-RNAs (miRNA) and mRNA expression demonstrated that miRNAs also play critical roles in regulating CD patients' gene expressions (Peck et al., 2015). Some of these circulating miRNAs were also thought to be potential prognosis markers for CD patients (Jensen et al., 2015). These data suggested that transcriptional and post-transcriptional regulation may play important roles in determining the etiology of CD.

Protein diversity determines the complexity of eukaryotic cellular processes. Alternative splicing (AS) of the precursor mRNA is one of the essential mechanisms to increase the protein diversity and regulate the intricate protein-RNA network (Chen and Manley, 2009; Wang et al., 2015). Nearly 95% of the total human multi-exon genes involve AS events (Black, 2003). Evidences indicates that AS plays crucial roles in not only oncogenic processes including proliferation, apoptosis, hypoxia, angiogenesis, immune escape and metastasis (David and Manley, 2010) (Oltean and Bates, 2014), but also in basic developmental processes and tissue identity (Baralle and Giudice, 2017). A pioneering study, which focused on alternative splicing of pre-mRNA in IBDs, has identified 47 splicing factors and 33 intron retention events that were dysregulated in mucosal tissue of patients (Häsler et al., 2011). However, due to the lack of

next generation sequencing technology, only 149 splicing factors and 145 intron retention events were screened in that pioneer study.

We reasoned that dysregulated AS could contribute to the pathogenesis of CD. Thus, we obtained public mRNA-seq data from NCBI GEO dataset (GSE66207) (Peck et al., 2015), which includes 33 colon tissue mRNA-seq data (20 are CD and 13 are healthy controls). Using Miso and related AS event analysis software (Katz et al., 2010a), we identified approximately 3,000 significant AS events in CD patients compared to the controls. 117 biological pathways are significantly enriched for these AS events, of which some are highly related to inflammation related responses. IRF1, STAT3, and MAOA were identified as the key relevant genes involved in these AS events. In a validation study, we also identified 64% genes of significant AS events overlapped with the results from the public dataset, indicating a comprehensive role of AS in CD. We believe our results may shed a light on understanding the mechanism of post-transcriptional regulation in CD progression.

Method

Sample Collection

A total of four patients with CD from Shengjing Hospital of China Medical University were collected between June 2020 and September 2020, and served as the experimental group. All eligible patients had an established diagnosis of CD according to endoscopic and histologic assessments. Colonic biopsy specimens were taken from the rectum, ulcer margin of sigmoid colon and inflamed portions. Four patients with normal distal colon confirmed by surgical pathology served as the control group. The study was approved by the institutional review board of Shengjing Hospital of China Medical University, and informed consent was obtained from each patient.

Data preparation

RNA-seq data was obtained from NCBI GEO dataset (GSE66207) (Peck et al., 2015). The sequencing was performed via ILLUMINA (Illumina HiSeq 2500) mean run: 26.98M spots. The mean size for the sequencing file includes 2.58G bases according to the SRX886282 project detail. This dataset included 20 Crohn's disease patients' and 13 normal colon tissues' raw sequence data and mRNA expression data (Patient info were included in the Table S4). Fastq files were aligned on human Hg19 genome by STAR-2.7.1a (Dobin et al., 2013) (Read length = 150, Average read depth = 50).. The indexed .bam file were generated by Samtools (1.10) (Li et al., 2009) for next step analysis. Raw files in validation the experiment were prepared in same method.

Identification of significant alternative splicing Events

Alternative splicing events were identified and quantified using MISO version 0.5.4 (Katz et al., 2010b) for 33 samples. A read length of 48 was applied and other parameters were in default in MISO. The level of alternative splicing events was defined as Percentage Spliced In (Psi). To get differentially expressed alternative splicing events between Crohn's disease patients and normal samples, we applied Wilcoxon test on all alternative splicing events with at least 3 patients and 1 normal sample. Significant events were defined as ones with p-values < 0.05. Same method was applied in transferability study part.

Pathway Analysis

Pathway analysis of the splicing and data was performed using Enrichr (Chen et al., 2013) (Kuleshov et al., 2016) with Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway library. Only the pathways with p-value < 0.05 were considered as related pathways. WebGestalt(Liao et al., 2019), SUMER (Savage et al., 2019) and Cytoscape (3.8.0) (Shannon et al., 2003) were applied in pathway analysis of transferability study part. Over-Representation Analysis (ORA) and geneontology biological process were applied as enrichment method and functional database respectively in WebGestalt. The enriched category with the gene sizes less than 5-kb and false discovery rate (FDR) above 0.05 was removed in WebGestalt results. The results of different alternative splicing types were input into SUMER to get pathway network. Finally, Cytoscape (3.8.0) was applied to modify color and text size of the network from SUMER.

Visualization

Sankey diagram was generated by Plotly in python. UpSet plots were generated by ComplexHeatmap (Gu et al., 2016) in R. Sashimi plots were generated by IGV (version 2.8.3) (Thorvaldsdóttir et al., 2013). In order to avoid random error, we randomly selected and merged RNA-seq alignment data from three Crohn's disease patients and three normal samples respectively. Other graphs were plotted using ggplot in R.

RNA-seq expression analysis

Pair-ended RNA-seq libraries for the transferability study were prepared using the NEBNext Ultra RNA with Poly-A selection kit and was sequenced on an Illumina Hi-Seq 4000 (Generry, Shanghai). Kallisto software was then used to quantify RNA-seq raw counts based on a pseudoalignment algorithm (Bray et al., 2016) (Read Length = 200, default; Mean read depth = 50). Differential gene expression was determined with log₂foldchange > 1.5 and P < 0.05 genes with >1 count per million. Any gene with a P-value greater than False Discovery Rate (FDR), after Benjamini–Hochberg correction for multi-testing, was deemed significantly differentially expressed under the test conditions as compared to the controls.

Result

Splicing Types

In the study of identifying alternative splicing events of the data from GSE66207, we found eight types of alternative splicing (AS) events including alternative 3'splice site (A3SS), alternative 5'splice site (A5SS), alternative first exon (AFE), alternative last exon (ALE), mutually exclusive exons (MXE), retrained introns (RI), skipped exon (SE), and Tandem 3' UTR. 2,980 significant alternative splicing events were discovered in 20 Crohn's disease patients (**Figure 1A**) (**Table S1**). Among these types, SE has the most significant alternative splicing events, but the percentage over total number of SE events is rather low. 4.1% of AFE events and 4% of Tandem 3'UTR events were recognized as significant events, indicating a higher level compared to the others (**Figure 1B**). The distribution and intersection of gene symbols among 8 events types for 2980 significantly different AS events were shown in **Figure 1C**

Pathway Analysis

The AS events which were related to Crohn's disease contribute to 117 different biological pathways (KEGG pathway) (**Figure 2**). Among these significantly enriched pathways, "Lysine degradation", "Sphingolipid metabolism", and "Proteoglycans in cancer", are the most enriched terms, which have the p-value 3×10^{-5} , 3.6×10^{-4} and 3.6×10^{-4} , respectively (**Table S2**). Most of the pathways were associated with only one splicing type. However, "Human cytomegalovirus infection" and "Lysin degradation" pathways were affected by four splicing types. We then combined related AS events pathway analysis results with the RNA-seq expression results. We found 8 inflammation related genes were regulated by CD related AS events. Among these 8 genes. Interferon Regulatory Factor 1 (IRF1), Signal Transducer And Activator Of Transcription 3 (STAT3) and Monoamine Oxidase A (MAOA), are identified to have the differentially expressed AS events with the lowest p-Value of Wilcoxon test and most significant differentially expression level, between the Crohn's disease and normal group (**Figure. 3A-C**). Three genes were all observed with SE splicing events. STAT3 was also observed with AFE splicing events. According to the pathway analysis, IRF1 and STAT3 are involved in prolactin signaling pathways which have a multitude of effects relating to immunoregulation and protection (Ben-Jonathan et al., 2006). Moreover, MAOA is related to tryptophan metabolism, which was also highly associated with inflammation, stress response, and microbiome homeostasis (Gostner et al., 2020). All of three genes in the different sample groups show significantly different RNA expression levels.

PCA analysis

We summarized 2,980 related AS events which occurred across all 33 samples with Percent Spliced In (PSI) values generated by MISO. To characterize the AS events between disease and normal sample, we performed a principal component analysis (PCA) on the MISO results of 2,980 related AS events. PC1-PC3 accounted for 33% of the variance (**Figure 4**). The biological differences between CD patients and normal samples were captured by the first, the second and third principal component (PC).

Splicing types in transferability study

To validate AS events that we identified in the GSE66207 datasets, we collected samples and performed RNA-seq on four colon tissues from patients with CD and four controls from Shengjing Hospital of China Medical University. Between June 2020 and September 2020, a total of 4 patients diagnosed on the basis of the standard clinical, endoscopic, and histological criteria of CD at the Gastroenterology Departments of the Shengjing Hospital of China Medical University were included and served as the validation group. The biopsies were obtained during endoscopy. Patients who had infectious colitis, indeterminate colitis, Behcet's disease, intestinal tuberculosis, and colorectal cancer were excluded. Colon biopsies that were obtained from healthy volunteers during endoscopy served as the control group. The clinical disease activity was assessed by the measurement of the Crohn's disease activity index for CD by a physician who was blinded to the colonoscopy appearance of the patient. The study was approved by the institutional review board of Shengjing Hospital of China Medical University, and informed consent was obtained from each patient (Table 1).

Table1 Clinical Characteristics of the Patients (A total of 4 patients with CD were included between June 2020 and September 2020. The clinical characteristics of all patients are shown in Table1. There were no significant differences with regard to age and sex between the experimental and control groups ($P > 0.05$).)

	patient1	patient2	patient3	patient4
Age (years)	16	22	30	23
Gender	Male	Female	Female	Male
Course of disease (years)	1.5	0.5	4	3
Age at diagnosis	15	22	29	21
Disease phenotype	B3	B1	B1p	B2p
Disease location	L3	L3	L3	L2
CDAI	160	170	410	240
Medication	Anti-TNF-a therapy	5-ASA and TNF-a therapy	Anti-TNF-a therapy	Anti-TNF-a therapy and Corticosteroids

Locations: L1 = terminal ileum; L2 = colon; L3= ileocolon. Behaviors: B1= nonstricturing;

B2 =stricturing, B3 =penetrating; P= perianal disease.

ASA, aminosalicylic acid.

Using the same analysis pipeline, we identified 1,715 significant AS events (**Table S3**) in this validation analysis. Interestingly, SE and AFE are still the most two type events and MXE is the least one, which is similar to the result from the public dataset (**Figure 5A**). But the percentage over the total number of all events types are relatively lower than the previous study (**Figure 5B**). In 1,715 significant AS events, AFE and ALE were also identified across multiple genes, which had the most intersections (**Figure 4C**).

Biological Process Pathway analysis in transferability study

We performed biological process pathway analysis in transferability study based on gene lists of 1,731 significant AS events. Since there are only less than 200 related genes in MXE, RI and Tandem 3' UTR events types, we didn't identify significant biological processes of these three types of events. In the other

five AS type events, we found several biological process networks were highly enriched using the WebGestalt analysis. Among these significantly enriched networks, “defense response”, “cell killing”, “response to cytokine” and their related modules were the most enriched clusters (**Figure 6**). These results suggested potential dysregulation of AS may play important roles in host-microbiome interactions in these patients with CD.

Comparison between the data from public dataset and our transferability study

Finally, we compared the results from the analysis of the public dataset and the results from the analysis of our own cohort. In terms of the genes that had significant AS events, we identified 64% genes that were identified to include significant AS events in transferability study were overlapped with the results we found in the public dataset (**Figure 7A**). Interestingly, IRF1 and STAT3 are still identified to have significant Tandem 3'UTR events and showed the significantly altered expression at mRNA level in the patients (**Figure 7B**). Finally, we collected 404 splicing factors from a large splicing factor database for expression analysis (Seiler et al., 2018). We excluded 34 factors that were barely expressed in the RNA-seq data (FPKM (Fragments Per Kilobase of transcript per Million mapped reads) < 1). As a result, we identified five splicing factors genes were significantly differentially expressed (p-adj-value < 0.05) in both dataset (**Table 2**). We also noticed that these splicing factors all had the same trend in terms of the expression level alteration. We speculated these splicing factors may be important in regulating the AS events in CD. In conclusion, these results suggested a strong similarity between our transferability study and the public dataset. We also speculated the difference from these analyses may be due to the different genetic background of these samples, as the first cohort was from the University of North Carolina (assumably a cohort of mixed ethnicity) the validation cohort consists of all Han Chinese. Besides, the average age of the first cohort was 45 years old, the average age of our cohort was 21 years old, indicating a significant difference on age between these two population. However, the sex ratio between the two cohort was similar (1:1 vs 0.7:1). These factors could also potentially affect the AS results.

Table 2. Transferability study RNA-seq splicing factor raw count fold change of 8 samples

Gene symbol	Description	GSE66207- Log2Foldchange CD vs Control	Validation experiment- Log2Foldchange CD vs Control
HSPA1A	Heat Shock Protein Family A (Hsp70) Member 1A	2.4950086	1.73285
RNF213	Ring Finger Protein 213	1.5825733	3.760895
MOV10	Mov10 RISC Complex RNA Helicase	1.5125286	1.304929
HSPA5	Heat Shock Protein Family A (Hsp70) Member 5	1. 3313641	3.431851
LSM10	LSM10, U7 Small Nuclear RNA Associated	1.2935081	1.631783

Discussion

Dysregulation of mRNA splicing can cause human disease, and more and more studies showed that targeting alternative splicing could lead to the development of novel therapeutics (Zhao, 2019). To date, accumulated data have already highlighted the crucial roles of AS in many human diseases, including cancer, neurodegenerative disorders, immune and infectious diseases (Wang and Cooper, 2007; Scotti and Swanson, 2016; Kim et al., 2018). Herein, we presented a comprehensive analysis of AS events in CD using a well-established public dataset and a validation dataset from sequencing a Chinese cohort. In our study, we found SE is the most significantly enriched AS event in CD. This suggested that the CD patients' transcriptome is much more complex and chaotic compared to the healthy controls, due to a more sophisticated mRNA isoform system. The PCA plot also demonstrated that even if only using different AS events as the principle components, we were still able to differentiate the CD group and control group, suggesting that the mRNA AS event itself, other than the expression level, could be a solid comprehensive signature for CD patients. All these results proved our hypothesis that dysregulated AS events may contribute to the etiology of CD.

Among the AS event related significantly enriched pathways, "Lysine degradation" pathway is the most enriched pathway. EZH1 and EZH2 are two genes involved in this pathway, which have been previously linked to the progression of IBD (Fernandez-Becker and Moss, 2009; Lou et al., 2019). They both have dysregulated SE events in CD patients. "Sphingolipids metabolism" is the most significantly enriched pathways in A3SS events. Interestingly, sphingolipids have recently recognized as mediators of

inflammation and potential therapeutic targets for IBDs (Sukocheva et al., 2020). Our results suggested that metabolism of sphingolipids in CD patients may differ from the healthy people due to AS, resulting in a dysregulated sphingolipid pool. Moreover, the three most dysregulated genes in inflammatory response related pathways were all previously linked to IBD. IRF1, as an interferon regulatory factor, has been correlated with CD activity index and CD endoscopic index of severity (Tang et al., 2014). In a cohort study, IRF1 has a 72% increase in gene expression among CD patients compared to the controls (Huff et al., 2012). STAT3, as a central component in immune response signal transduction, if has gain-of-function mutation, will cause multi-organ early auto-immune diseases (Milner et al., 2015). Notably, STAT3 has also been indicated as one of the crucial targets for treating IBDs, although activation of STAT3 is likely to occur in both innate and acquired cell types (Sugimoto, 2008). MAOA, as a monoamine oxidase, though is not directly associated with IBDs, is highly related with inducing oxidative stress in obese people who are suffering chronic inflammation (Sturza et al., 2019). Recently, dysregulation of MAOA has also been linked to specific microbe alteration (Xie et al., 2020). In conclusion, both top AS related significantly enriched pathways and significantly dysregulated genes are highly related with IBD and CD. Our results provided new insights that post-transcriptional regulation could also contribute to IBD via regulating these key pathways and genes.

Although our study provided a comprehensive AS analysis on colon tissues from CD patients, and also have identified nearly 3,000 unique AS events in CD patients, our studies still have certain limitations. In the first place, we still need to validate these AS events experimentally. Moreover, whether these AS events affect protein diversity also requires experimental validation. The validation studies should include examining the protein level of these AS regulated genes, and the detecting the downstream effect on certain signaling transduction pathways. Finally, even if we have detected several splicing factors were dysregulated in CD patients, a RIP-seq or CLIP-seq study should be performed to further detect the responsible splicing factors, RNA binding proteins for CD related AS events. Nonetheless, our study demonstrated a new potential pathophysiology of CDs which is regulated by AS. Our study also presented the first landscape of AS in Chinses patients. Our data also suggested that drugs that target AS related genes and splicing factors should be considered for further screening for CD patients.

Conclusion

Our study presented a landscape of alternative splicing (AS) in CD, and provided the first AS related transcriptional analysis in Chinese cohort. Integrated analysis of two datasets revealed that AS may play a crucial role in determining the pathogenesis of CD.

Abbreviations

CD: Crohn's disease; AS: alternative splicing; IBD: inflammatory bowel disease; miRNAs: Micro-RNA; KEGG: Kyoto Encyclopedia of Genes and Genomes; FDR: False Discovery Rate; 3' UTR: three prime untranslated region.

Declarations

- Ethics approval and consent to participate
- The study was approved by the institutional review board of Shengjing Hospital of China Medical University, and informed consent was obtained from each patient.
- Consent for publication
- All samples were collected with written informed parental consent for publication.
- Availability of data and material
- Data are available from the corresponding author on request. Please contact author for data requests.
- Competing interests
- The authors declare no competing financial interests.
- Funding
- This work was supported partly by the Science and Technology Program of Liaoning Province (No. 2019-BS-140).
- Authors' contributions
- Y. Tan designed research and collected the patients' samples; D. Li analyzed data; D. Li and Y. Tan wrote the paper.
- Acknowledgements
- We appreciate the assistance from Intanx Life (Shanghai) Co. Ltd. in data processing and consulting.

References

- Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology* 18, 437.
- Ben-Jonathan, N., Hugo, E.R., Brandebourg, T.D., and Lapensee, C.R. (2006). Focus on prolactin as a metabolic hormone. *Trends Endocrinol Metab* 17, 110-116.
- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry* 72, 291-336.
- Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* 34, 525-527.
- Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics* 14, 128.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature reviews Molecular cell biology* 10, 741-754.

David, C.J., and Manley, J.L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* 24, 2343-2364.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21.

Fernandez-Becker, N.Q., and Moss, A.C. (2009). In silico analysis of T-bet activity in peripheral blood mononuclear cells in patients with inflammatory bowel disease (IBD). *In Silico Biol* 9, 355-363.

Gostner, J.M., Geisler, S., Stonig, M., Mair, L., Sperner-Unterweger, B., and Fuchs, D. (2020). Tryptophan Metabolism and Related Pathways in Psychoneuroimmunology: The Impact of Nutrition and Lifestyle. *Neuropsychobiology* 79, 89-99.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847-2849.

Häsler, R., Kerick, M., Mah, N., Hultschig, C., Richter, G., Bretz, F., Sina, C., Lehrach, H., Nietfeld, W., Schreiber, S., and Rosenstiel, P. (2011). Alterations of pre-mRNA splicing in human inflammatory bowel disease. *Eur J Cell Biol* 90, 603-611.

Hong, S.N., Joung, J.-G., Bae, J.S., Lee, C.S., Koo, J.S., Park, S.J., Im, J.P., Kim, Y.S., Kim, J.W., Park, W.Y., and Kim, Y.-H. (2017). RNA-seq Reveals Transcriptomic Differences in Inflamed and Noninflamed Intestinal Mucosa of Crohn's Disease Patients Compared with Normal Mucosa of Healthy Controls. *Inflammatory Bowel Diseases* 23, 1098-1108.

Hong, S.N., Park, C., Park, S.J., Lee, C.K., Ye, B.D., Kim, Y.S., Lee, S., Chae, J., Kim, J.-I., and Kim, Y.-H. (2016). Deep resequencing of 131 Crohn's disease associated genes in pooled DNA confirmed three reported variants and identified eight novel variants. *Gut* 65, 788-796.

Huff, C.D., Witherspoon, D.J., Zhang, Y., Gatenbee, C., Denson, L.A., Kugathasan, S., Hakonarson, H., Whiting, A., Davis, C.T., Wu, W., Xing, J., Watkins, W.S., Bamshad, M.J., Bradfield, J.P., Bulayeva, K., Simonson, T.S., Jorde, L.B., and Guthery, S.L. (2012). Crohn's disease and genetic hitchhiking at IBD5. *Mol Biol Evol* 29, 101-111.

Jensen, M.D., Andersen, R.F., Christensen, H., Nathan, T., Kjeldsen, J., and Madsen, J.S. (2015). Circulating microRNAs as biomarkers of adult Crohn's disease. *Eur J Gastroenterol Hepatol* 27, 1038-1044.

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010a). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature methods* 7, 1009-1015.

- Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010b). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* 7, 1009-1015.
- Kim, H.K., Pham, M.H.C., Ko, K.S., Rhee, B.D., and Han, J. (2018). Alternative splicing isoforms in health and disease. *Pflügers Archiv-European Journal of Physiology* 470, 995-1016.
- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., and Lachmann, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* 44, W90-W97.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic acids research* 47, W199-W205.
- Lou, X., Zhu, H., Ning, L., Li, C., Li, S., Du, H., Zhou, X., and Xu, G. (2019). EZH2 Regulates Intestinal Inflammation and Necroptosis Through the JNK Signaling Pathway in Intestinal Epithelial Cells. *Dig Dis Sci* 64, 3518-3527.
- Milner, J.D., Vogel, T.P., Forbes, L., Ma, C.A., Stray-Pedersen, A., Niemela, J.E., Lyons, J.J., Engelhardt, K.R., Zhang, Y., and Topcagic, N. (2015). Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations. *Blood* 125, 591-599.
- Momozawa, Y., Dmitrieva, J., Théâtre, E., Deffontaine, V., Rahmouni, S., Charlotteaux, B., Crins, F., Docampo, E., Elansary, M., Gori, A.-S., Lecut, C., Mariman, R., Mni, M., Oury, C., Altukhov, I., Alexeev, D., Aulchenko, Y., Amininejad, L., Bouma, G., Hoentjen, F., Löwenberg, M., Oldenburg, B., Pierik, M.J., Vander Meulen-De Jong, A.E., Janneke Van Der Woude, C., Visschedijk, M.C., Abraham, C., Achkar, J.-P., Ahmad, T., Ananthakrishnan, A.N., Andersen, V., Anderson, C.A., Andrews, J.M., Annese, V., Aumais, G., Baidoo, L., Baldassano, R.N., Bampton, P.A., Barclay, M., Barrett, J.C., Bayless, T.M., Bethge, J., Bitton, A., Boucher, G., Brand, S., Brandt, B., Brant, S.R., Büning, C., Chew, A., Cho, J.H., Cleynen, I., Cohain, A., Croft, A., Daly, M.J., D'amato, M., Danese, S., Jong, D.D., Denapiene, G., Denson, L.A., Devaney, K.L., Dewit, O., D'inca, R., Dubinsky, M., Duerr, R.H., Edwards, C., Ellinghaus, D., Essers, J., Ferguson, L.R., Festen, E.A., Fleshner, P., Florin, T., Franke, A., Fransen, K., Gearry, R., Gieger, C., Glas, J., Goyette, P., Green, T., Griffiths, A.M., Guthery, S.L., Hakonarson, H., Halfvarson, J., Hanigan, K., Haritunians, T., Hart, A., Hawkey, C., Hayward, N.K., Hedl, M., Henderson, P., Hu, X., Huang, H., Hui, K.Y., Imielinski, M., Ippoliti, A., Jonaitis, L., Jostins, L., Karlsen, T.H., Kennedy, N.A., Khan, M.A., Kiudelis, G., et al. (2018). IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nature Communications* 9, 2427.
- Oltean, S., and Bates, D.O. (2014). Hallmarks of alternative splicing in cancer. *Oncogene* 33, 5311-5318.
- Peck, B.C.E., Weiser, M., Lee, S.E., Gipson, G.R., Iyer, V.B., Sartor, R.B., Herfarth, H.H., Long, M.D., Hansen, J.J., Isaacs, K.L., Trembath, D.G., Rahbar, R., Sadiq, T.S., Furey, T.S., Sethupathy, P., and Sheikh, S.Z. (2015).

MicroRNAs Classify Different Disease Behavior Phenotypes of Crohn's Disease and May Have Prognostic Utility. *Inflammatory Bowel Diseases* 21, 2178-2187.

Savage, S.R., Shi, Z., Liao, Y., and Zhang, B. (2019). Graph algorithms for condensing and consolidating gene set analysis results. *Molecular & Cellular Proteomics* 18, S141-S152.

Scotti, M.M., and Swanson, M.S. (2016). RNA mis-splicing in disease. *Nature Reviews Genetics* 17, 19.

Seiler, M., Peng, S., Agrawal, A.A., Palacino, J., Teng, T., Zhu, P., Smith, P.G., Buonamici, S., and Yu, L. (2018). Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep* 23, 282-296.e284.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13, 2498-2504.

Simmons, A. (2010). Genes, viruses and microbes. *Nature* 466, 699-700.

Sturza, A., Olariu, S., Ionică, M., Duicu, O.M., Văduva, A.O., Boia, E., Muntean, D.M., and Popoiu, C.M. (2019). Monoamine oxidase is a source of oxidative stress in obese patients with chronic inflammation (1). *Can J Physiol Pharmacol* 97, 844-849.

Sugimoto, K. (2008). Role of STAT3 in inflammatory bowel disease. *World journal of gastroenterology: WJG* 14, 5110.

Sukocheva, O.A., Lukina, E., McGowan, E., and Bishayee, A. (2020). Sphingolipids as mediators of inflammation and novel therapeutic target in inflammatory bowel disease. *Adv Protein Chem Struct Biol* 120, 123-158.

Tang, R., Yang, G., Zhang, S., Wu, C., and Chen, M. (2014). Opposite effects of interferon regulatory factor 1 and osteopontin on the apoptosis of epithelial cells induced by TNF- α in inflammatory bowel disease. *Inflamm Bowel Dis* 20, 1950-1961.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 14, 178-192.

Van Beelen Granlund, A., Flatberg, A., Østvik, A.E., Drozdov, I., Gustafsson, B.I., Kidd, M., Beisvag, V., Torp, S.H., Waldum, H.L., and Martinsen, T.C. (2013). Whole genome gene expression meta-analysis of inflammatory bowel disease colon mucosa demonstrates lack of major differences between Crohn's disease and ulcerative colitis. *PloS one* 8.

Wang, G.-S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics* 8, 749-761.

Wang, Y., Liu, J., Huang, B., Xu, Y.M., Li, J., Huang, L.F., Lin, J., Zhang, J., Min, Q.H., and Yang, W.M. (2015). Mechanism of alternative splicing and its regulation. *Biomedical reports* 3, 152-158.

Xie, Y., Wang, C., Zhao, D., Wang, C., and Li, C. (2020). Dietary Proteins Regulate Serotonin Biosynthesis and Catabolism by Specific Gut Microbes. *Journal of Agricultural and Food Chemistry*.

Zhao, S. (2019). Alternative splicing, RNA-seq and drug discovery. *Drug Discovery Today* 24, 1258-1267.

Figures

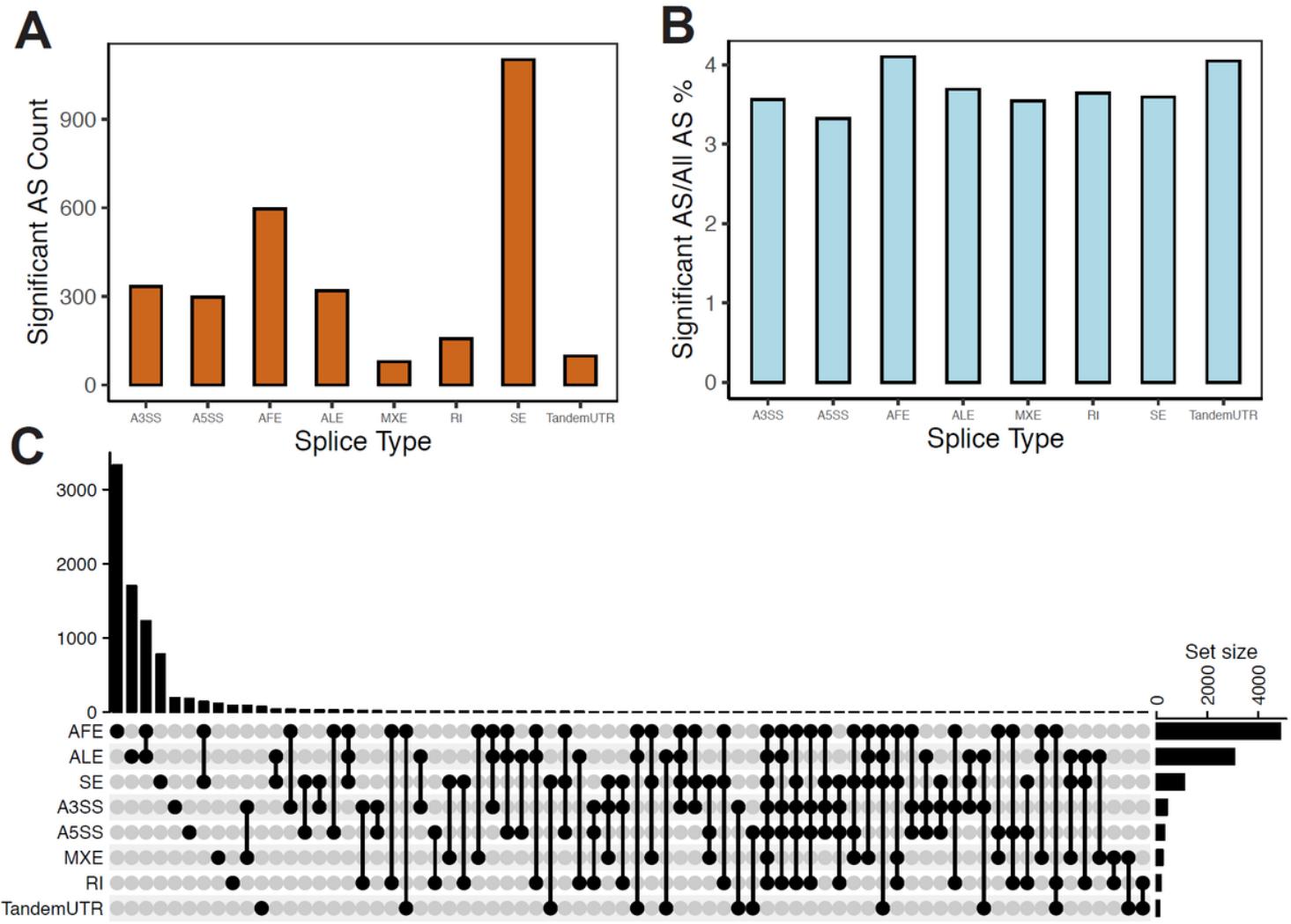


Figure 1

Alternative splicing identification in CD patients and normal colon samples. A. Significant AS event counts in different AS types between Crohn's disease patients and normal colon samples. B. The percentage of significant AS events over all AS events in different AS types. C. Distribution of AS types and intersection between gene symbols of 2,980 related AS events.

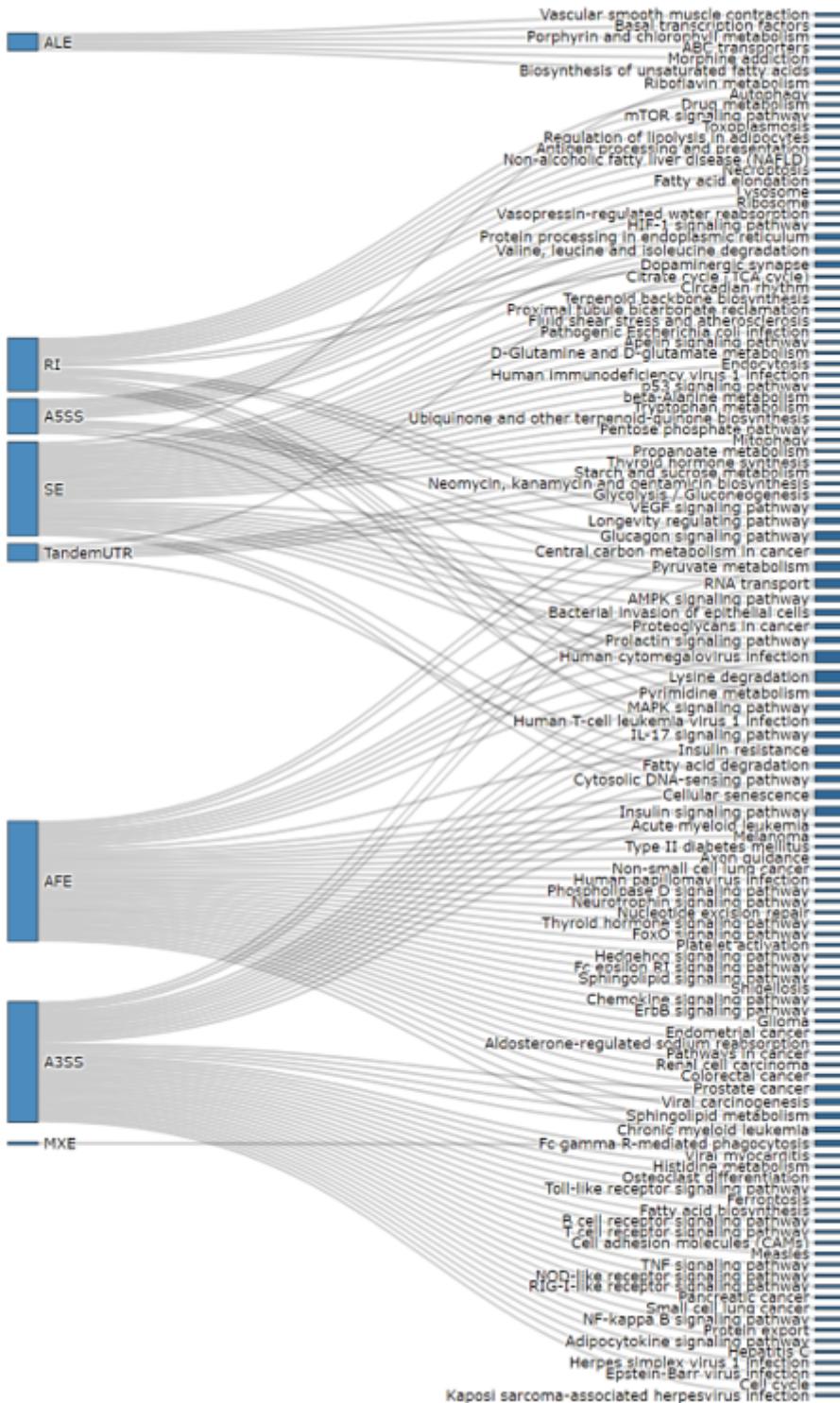


Figure 2

Pathway Analysis. The biological pathways associated with the related AS events in different splice type.

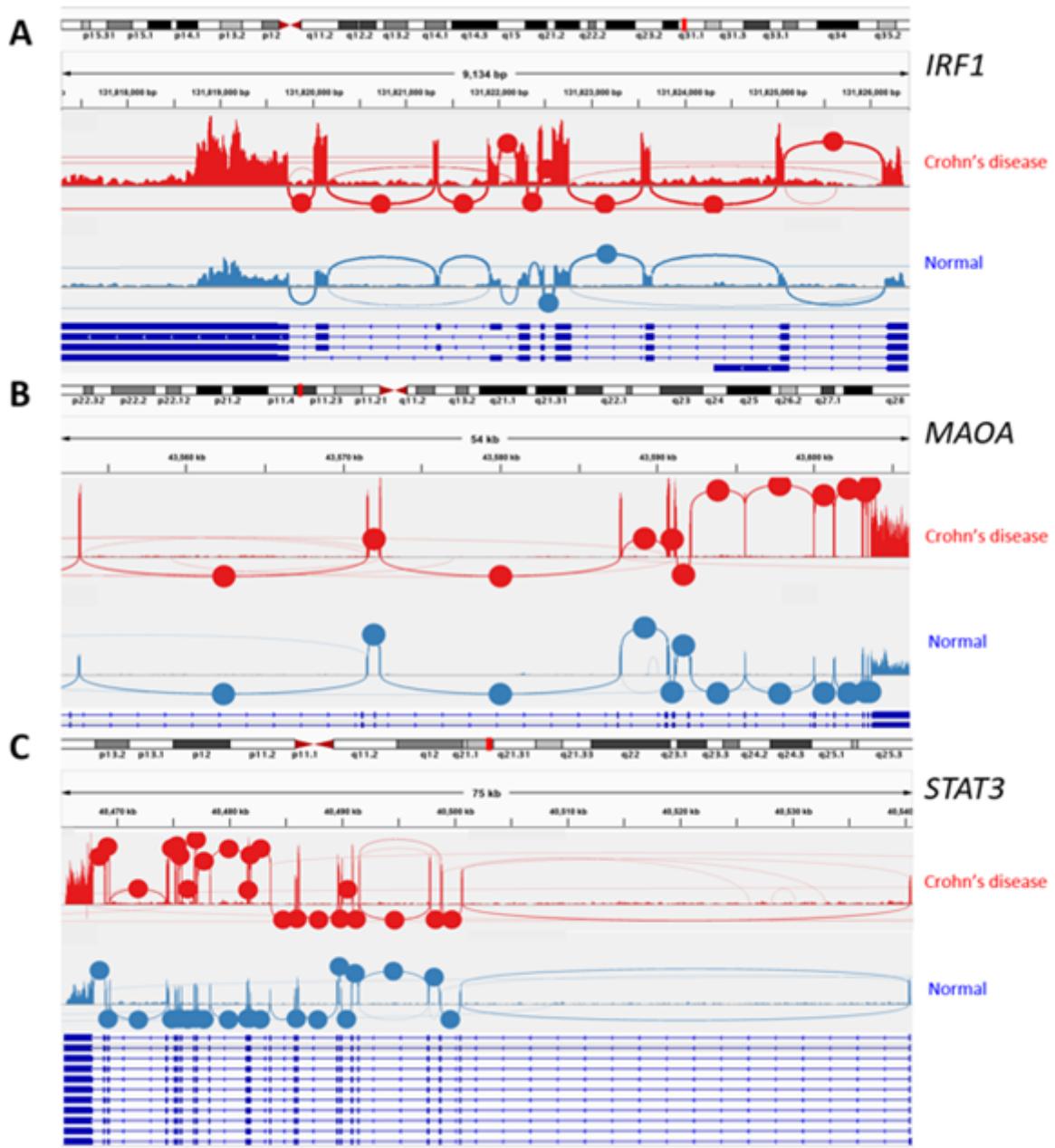


Figure 3

IGV Sashimi Plot of three related genes (*IRF1*, *MAOA* and *STAT3*) expression and potential splicing patterns (circle = junction coverage, line = potential splicing event).

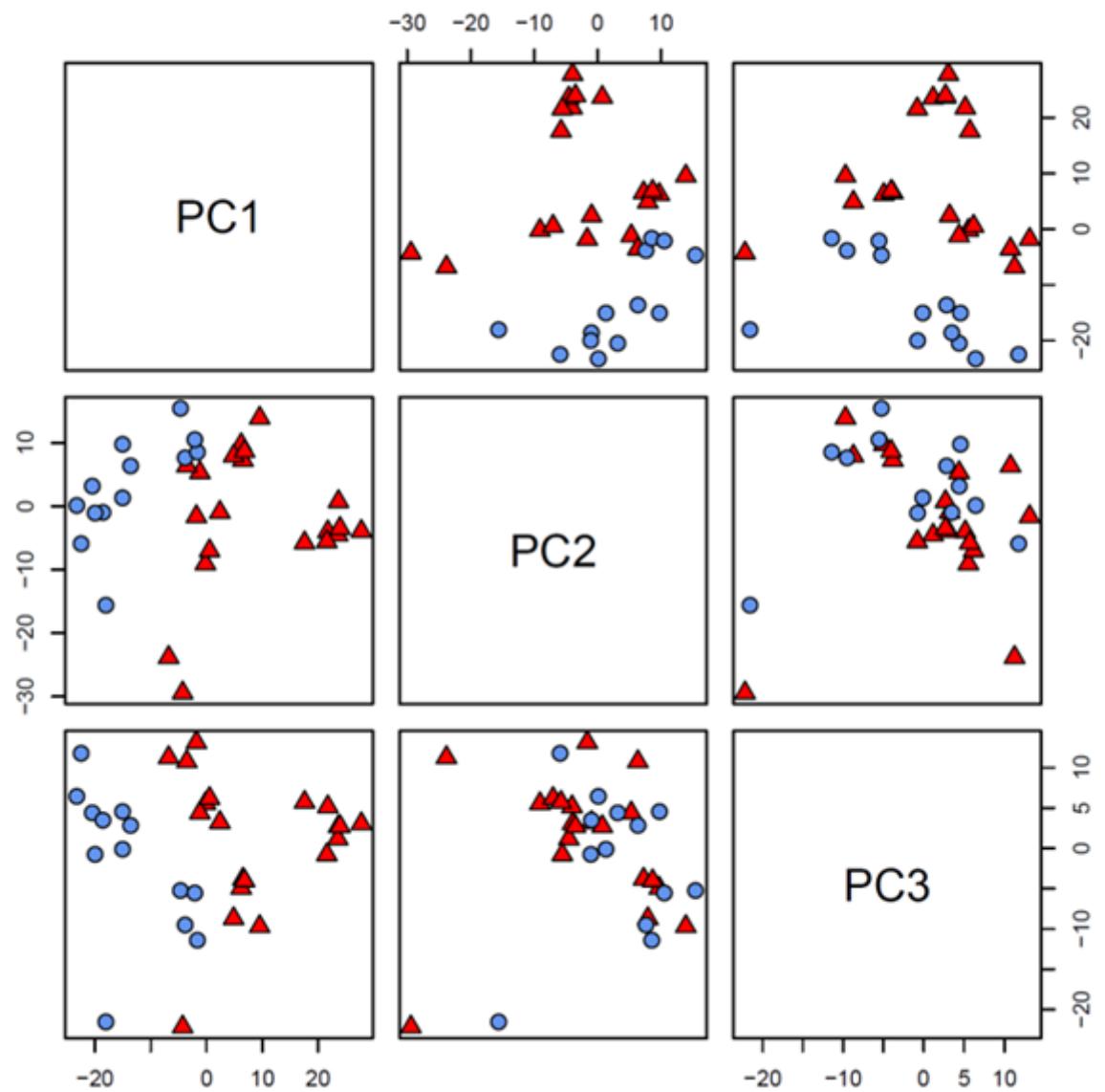


Figure 4

PCA plot. Scatterplot pairs of first three principal components for Percent Spliced In (PSI) values across 33 samples. The red triangles represent CD patients and the blue circle represent normal samples.

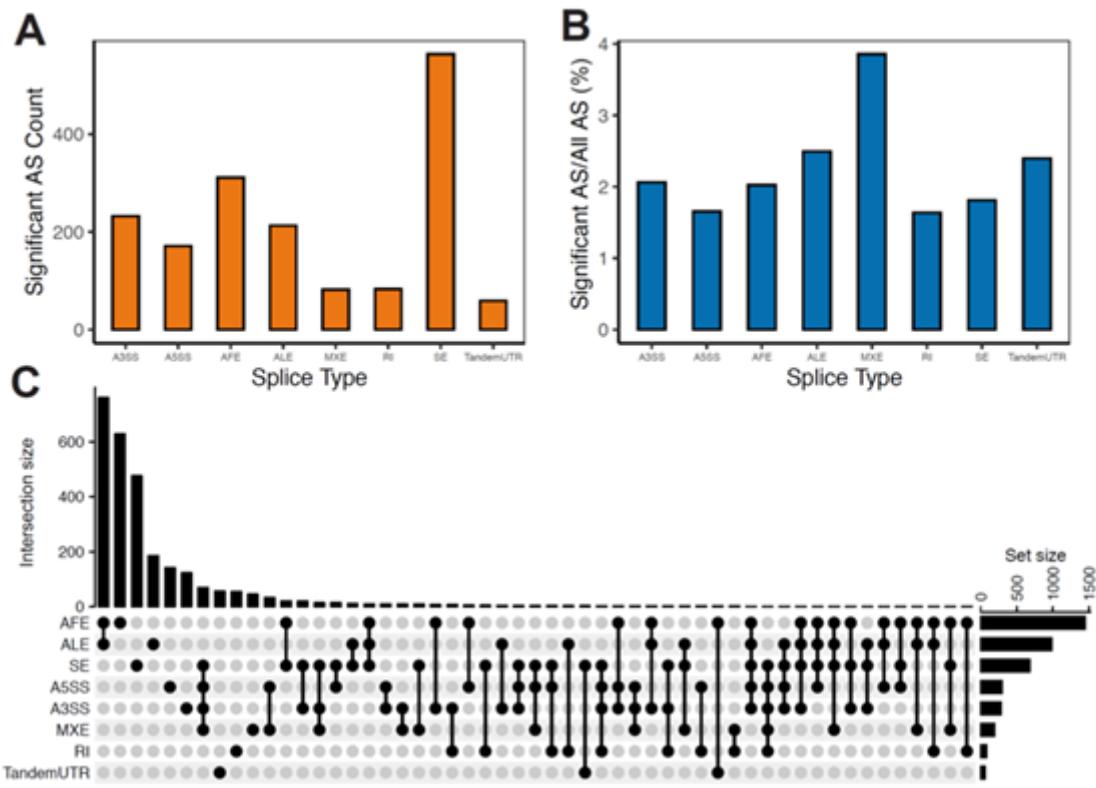


Figure 5

Alternative splicing identification in CD patients and normal samples in transferability study. A. Significant AS event counts in different AS types between CD patients and normal samples. B. The percentage of significant AS events over all AS events in different AS types. C. Distribution of AS types and intersection between gene symbols of 1,715 related AS events.



Figure 6

Enriched biological process networks which were identified in our validation study.

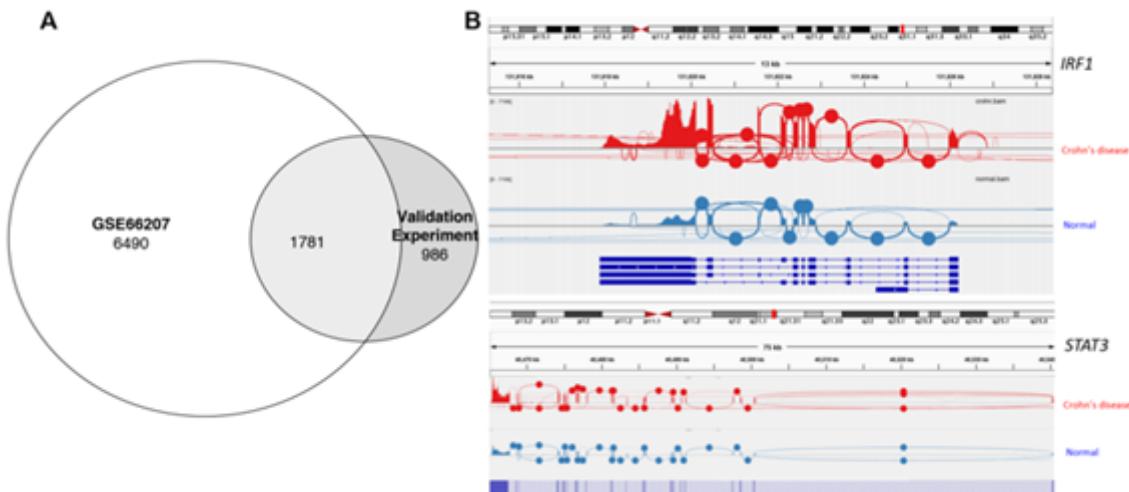


Figure 7

Linkage between the results from public dataset and the results from our transferability study. A. Venn diagram showed the 1781 genes that were identified to include significant AS events in transferability study were overlapped with the results from GSE66207. B. IGV Sashimi plot showed IRF1 and STAT3 also

had significant tandem 3'UTR events as well as significantly altered expression in patients with CD compared to the controls.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)