Disrupted auto-regulation of the spliceosomal gene SNRPB causes cerebro–costo–mandibular syndrome

Danielle C. Lynch1, Timothée Revil2, Jeremy Schwartzentruber3, Elizabeth J. Bhoj4, A. Micheil Innes1,5, Ryan E. Lamont1,5, Edmond G. Lemire6, Bernard N. Chodirker7,8, Juliet P. Taylor9, Elaine H. Zackai4, D. Ross McLeod1,5, Edwin P. Kirk10,11, Julie Hoover-Fong12, Leah Fleming13, Ravi Savarirayan14, Care4Rare Canada,1 Jacek Majewski2,3, Loydie A. Jerome-Majewska2,15, Jillian S. Parboosingh1,5, & Francois P. Bernier1,5,*

Elucidating the function of highly conserved regulatory sequences is a significant challenge in genomics today. Certain intragenic highly conserved elements have been associated with regulating levels of core components of the spliceosome and alternative splicing of downstream genes. Here we identify mutations in one such element, a regulatory alternative exon of SNRPB as the cause of cerebro–costo–mandibular syndrome. This exon contains a premature termination codon that triggers nonsense-mediated mRNA decay when included in the transcript. These mutations cause increased inclusion of the alternative exon and decreased overall expression of SNRPB. We provide evidence for the functional importance of this conserved intragenic element in the regulation of alternative splicing and development, and suggest that the evolution of such a regulatory mechanism has contributed to the complexity of mammalian development.
Although only 1.5% of the human genome consists of regions that are translated into proteins, a higher proportion (5–7%) has been shown to be under evolutionary constraint\textsuperscript{1,2}. These non-coding conserved elements (NCEs) have been subclassified by somewhat arbitrary length and conservation criteria, and include ultraconserved\textsuperscript{3} and highly conserved\textsuperscript{4} elements. The finding that many NCEs exhibit a higher level of conservation\textsuperscript{5} and constraint\textsuperscript{6} than protein-coding sequences initially perplexed the genomics community. The elucidation that some NCEs have functional roles as long range enhancers of flanking genes, splicing regulators, functional co-activators\textsuperscript{7–9} and their frequent association with developmental genes with the potential to regulate spatiotemporal expression\textsuperscript{7,10}, imply a largely regulatory role. The evolution of these complex regulatory networks may therefore have underpinned the emergence of our organismal complexity\textsuperscript{6,11}.

Evidence continues to emerge of a critical relationship between NCEs and alternative splicing (AS), the mechanism by which over 95% of human multi-exon genes create additional protein diversity\textsuperscript{12,13}. Intragenic NCEs are preferentially associated with genes involved in pre-mRNA splicing\textsuperscript{3} and are also often involved in the regulation of the expression of this class of genes by coupling AS with nonsense-mediated decay (NMD)\textsuperscript{14–16}. Many genes involved in pre-mRNA splicing have ultra and highly conserved introns in regions that are translated into proteins, a higher proportion (5–7%) has been shown to be under evolutionary constraint\textsuperscript{1,2}. These non-coding conserved elements (NCEs) have been subclassified by somewhat arbitrary length and conservation criteria, and include ultraconserved\textsuperscript{3} and highly conserved\textsuperscript{4} elements. The finding that many NCEs exhibit a higher level of conservation\textsuperscript{5} and constraint\textsuperscript{6} than protein-coding sequences initially perplexed the genomics community. The elucidation that some NCEs have functional roles as long range enhancers of flanking genes, splicing regulators, functional co-activators\textsuperscript{7–9} and their frequent association with developmental genes with the potential to regulate spatiotemporal expression\textsuperscript{7,10}, imply a largely regulatory role. The evolution of these complex regulatory networks may therefore have underpinned the emergence of our organismal complexity\textsuperscript{6,11}.

Here we present mutations in a highly conserved, alternative PTC-containing exon of the small nuclear ribonucleoprotein polypeptides B and B1 (SNRPB) gene (Fig. 1) as the cause of cerebro–costo–mandibular syndrome (CCMS), a human multiple malformation disorder characterized by posterior rib gaps and Pierre Robin sequence (micrognathia, glossoptosis and cleft palate). This finding provides biological evidence of a direct link between conserved genomic elements, regulation of AS and human development, and therefore novel insight in the regulatory and developmental role of NCEs.

Results
A combination of whole-exome sequencing and Sanger sequencing was used to identify causative mutations in a cohort of 10 unrelated families with CCMS, a rare genetic disorder characterized by micrognathia and posterior rib gaps\textsuperscript{19} (Supplementary Fig. 1). All patients have typical features of CCMS except one patient (Family D) who had a more severe disease (Supplementary Methods, Supplementary Fig. 2 and Supplementary Table 1). Nine of the 10 patients had heterozygous regulatory mutations in SNRPB. Overall, six distinct, novel mutations in SNRPB were identified. Five mutations are within the alternative PTC-containing exon (chr20g:2447388_2447961) of SNRPB. These mutations cluster at the 5’ and 3’ ends of this exon within areas of high conservation (Fig. 1c). A single patient had a 5’ untranslated region (UTR) mutation, which is predicted to introduce an out-of-frame translation initiation site (TIS) leading to a stop codon after 25 amino acids (Fig. 1b). In the SNRPB-positive families, mutation analysis confirms that CCMS is an autosomal dominant disorder. We observed a high rate of \textit{de novo} mutations and two instances of non-penetrance. One individual with classic CCMS was negative for sequence or copy-number variants in the coding and UTRs of SNRPB.

SNRPB encodes the protein isoforms SmB and SmB’\textsuperscript{17}, which are core components of the U1, U2, U4/U6 and U5 small ribonucleoprotein (snRNP)\textsuperscript{20} subunits of the major spliceosome. The highly conserved alternative exon within the second of six introns in SNRPB contains a PTC and has been shown to auto-regulate SNRPB levels through NMD\textsuperscript{18}. The alternate exon, which has a sub-optimal 5’ splice site, is less frequently included when U1 snRNP levels are low as a result of SM/B’ deple\textsuperscript{18}. Conversely, it is more frequently included with SmB/B’ overexpression\textsuperscript{16}. We hypothesized that the mutations identified within this exon would alter the homeostatic balance between the coding full-length mRNA and alternative exon-containing transcripts targeted for degradation. Thus, we determined the effect of two of the alternative exon mutations using a splicing reporter minigene assay\textsuperscript{21}. In the presence of the wild-type exon, 23% of all transcripts include this alternative exon, while introduction of either the chr20g:2447951C>G or chr20g:2447847G>T mutation shifts the proportion to 78% and 80%, respectively (Fig. 2a,b). Inclusion of the alternative PTC-containing exon was also assessed by quantitative reverse transcription PCR (qRT–PCR) in patient fibroblasts with the chr20g:2447951C>G, chr20g:2449752C>G, and chr20g:2447847T>G mutations. Expression of the PTC-containing transcript increased, whereas overall expression of SNRPB decreased compared with control cells (Fig. 2c,d).

![Figure 1: SNRPB mutations in CCMS.](image-url)
Discussion

Collectively, these results implicate the deregulation of SNRPB expression as the main disease mechanism for CCMS. Mutations in the alternative PTC-containing exon cluster at two sites, which overlap with known exonic splicing silencers (ESSs)\(^{22}\). In an experiment by Saltzman et al.\(^{18}\), deletion of both of these regions resulted in increased inclusion of the alternative exon in HeLa cells. Our results support the functional significance of these ESSs, which are perfectly conserved across placental mammals (Fig. 3 and Supplementary Fig. 3), and suggest that the identified mutations weaken their silencing function. This would lead to the observed increase in the inclusion of this exon in CCMS, which is presumably the cause of the decreased overall SNRPB expression seen in patient cells (Fig. 4).

The mutations identified in the alternative exon appear to cause a reduction in the amount of SmB/B\(^+\) that is consistent with a hypomorphic, but not a null, allele. qRT–PCR experiments in three patients show a narrow range of total SNRPB expression (0.53–0.66 relative to controls). In the minigene experiment, exclusion of the alternative exon was not eliminated in mutant transcripts, but occurred 20–22% of the time. We also have evidence that null alleles might result in a more severe phenotype as one patient without an alternative exon mutation has a 5\(^{\prime}\) UTR mutation predicted to result in a null allele causing haploinsufficiency (Supplementary Discussion). This patient’s phenotype was more severe than the remainder of the cohort, with only five pairs of poorly ossified ribs, a poorly ossified spine, cystic hygroma and multiple pterygia (Supplementary Table 1). Since none of the other patients carry truncating mutations in the gene (which would be much more likely to occur by chance than point mutations at two specific loci), and truncating mutations in or deletions encompassing SNRPB have not been reported\(^{23,24}\), we suggest that SNRPB haploinsufficiency may cause a more severe and likely lethal phenotype that is distinct from classic CCMS.

CCMS joins a growing list of developmental disorders caused by mutations in core spliceosomal genes. Of particular interest are those with an overlapping craniofacial phenotype, such as Nager syndrome and the EFTUD2-related disorders\(^{25–27}\). Interestingly, all of the above are caused by dominant mutations that are predicted to reduce expression of a component of the major spliceosome. It is known that the abundance of the spliceosomal machinery influences AS\(^{28,29}\). In the case of SNRPB, RNAseq experiments have shown that specific AS exons are more sensitive to changes in SmB/B\(^+\) levels\(^{18}\). Among genes containing such exons, nucleic acid binding and RNA processing genes are over-represented. The SNRPB mutations presented here are therefore predicted to cross-regulate AS and expression of downstream
Higher levels of these proteins then favour inclusion of the alternative exon, by an unknown mechanism, leading to NMD and a reduction of SmB/SmB′ protein levels. In alleles mutated in CCMS patients, the binding of repressor proteins is thought to be abolished or reduced due to the mutations present in the regulatory sequences. This leads to continued inclusion of the alternative exon, and reduced SmB/SmB′ protein levels due to NMD. A model of disrupted SNRPB regulation in CCMS. Unknown repressor proteins (red circles) bind the ESS regulatory sequences (red squares) in the alternatively spliced exon (in blue) of SNRPB. Their binding leads to exclusion of this alternative exon, and thus an increase of SmB/SmB′ protein levels. Higher levels of these proteins then favour inclusion of the alternative exon, by an unknown mechanism, leading to NMD and a reduction of SmB/SmB′ protein levels. In alleles mutated in CCMS patients, the binding of repressor proteins is thought to be abolished or reduced due to the mutations present in the regulatory sequences. This leads to continued inclusion of the alternative exon, and reduced SmB/SmB′ protein levels due to NMD.

Methods

Patients. A cohort of 10 CCMS families was assembled through the Finding Of Rare disease Genes (FORGE) Canada Consortium (now called Care4Rare). All patients provided informed consent, and the study was approved by and complies with the ethical regulations of the institutional review board at the University of Calgary. An experienced clinical geneticist was responsible for each diagnosis of CCMS. Exclusion criteria included absence of micrognathia and posterior rib gaps. Other variable features include scoliosis, short stature, conductive hearing loss and congenital heart defects. Although intellectual disability is reported to be a common feature of CCMS, this was not prevalent in our cohort (Supplementary Fig. 2 and Supplementary Table 1). Family A had a sibling recurrence with unaffected parents, families E and F had parent–child transmission, and the seven remaining cases were sporadic (Supplementary Fig. 1).

Exome sequencing. DNA was extracted from whole blood. Exome sequencing was performed for six unrelated cases and seven family members at the McGill University and Genome Québec Innovation Centre. The SureSelect 50 Mb Human All Exon kit (Agilent) was used for exon capture; v3 was used for families A, B and C, and v5 was used for families D, E and F. Captured regions were sequenced on a HiSeq 2000 sequencer (Illumina) with 100 bp paired-end reads. Reads were aligned to the hg19/GRCh37 human reference sequence using the Burrows-Wheeler Aligner and indel realignment was done with GATK. Duplicate reads were then marked using Picard and indel realignment was done with GATK. Duplicate reads were then marked using Picard and indel realignment was done with GATK. Duplicate reads were then marked using Picard and indel realignment was done with GATK. Duplicate reads were then marked using Picard and indel realignment was done with GATK.

Our study highlights the importance of accurate AS in development, alludes to the broad network of splicing regulation, and demonstrates the regulatory and developmental importance of a highly conserved regulatory element. The alternative exon of SNRPB has high conservation at the nucleotide level throughout placental mammals (average GERP score 4.08), although to a lesser extent than the ultra-conserved elements (Supplementary Fig. 3). In general, shorter human conserved elements are conserved among mammals, but not with other species. It has been suggested that evolution of these elements is ongoing in vertebrates, and that specific specializations may reflect clade-specific adaptive regulatory changes. It is then possible that auto-regulation of SNRPB has evolved in mammals with the function of guiding specific cellular and developmental processes. Broadly, we therefore speculate that NCEs may have a significant role in regulating the phenotypic variation on which natural selection acts to drive the evolution of complex and highly integrated traits.
1,000 genomes or NHBLI exome data sets, and manually reviewed remaining candidates. For family D a de novo 5’ UTR variant was seen that introduced a potential out-of-frame TIS in SNRPB. We used TIS miner (http://dnasminter.bio.nus.edu.sg/Tis.html) to predict the effect of this variant.

Sanger sequencing. For all individuals in the cohort, Sanger sequencing of the alternative exon including the flanking intronic regions was performed. For patient D II-1, the 5’ UTR was sequenced to confirm the presence of the variant identified by exome sequencing. For patient G II-3, the coding regions, including the flanking intronic sequences, and the UTRs of SNRPB were sequenced. Primers were designed with Oligo 6 (Molecular Biology Insights). Sequences can be found in Supplementary Table 5. These, along with the miniSmB (NEB) by incubation at 25°C.

Copy-number variant analysis by dPCR. dPCR of all exons of SNRPB was used to search for copy-number variants in patient G II-3. One microtiter of 5 ng ml−1 DNA was used in a 20-µl reaction with 1 µl of 10 µm primer mix, 10 µl of SYBR Green (Life Technologies) and 8 µl of H2O. Primer sequences can be found in Supplementary Table 3. Reactions were run on a 7900HT Fast Real-Time PCR System (Applied Biosystems). Cycling conditions were as follows: 96°C for 5:00, 35 cycles of 96°C for 0:30, 58°C for 0:30 and 72°C for 0:30, and a final elongation step at 72°C for 0:00. An amount of 5 µl was analysed on a 1% agarose gel. A quantity of 1.2 µl of 1/20 dilution of the PCR product was purified in a reaction with 1 µl ExoSAP-IT (Affymetrix) and 3 µl H2O. The product of this reaction was added to a sequencing reaction with 2.2 µl HiDi-1.875 µl of 5 × sequencing buffer, 0.5 µl primer and 0.25 µl BigDye Terminator v1.1 (Life Technologies). Unincorporated nucleotides were removed from the sequencing reaction by passage through a Sephadex column. The products were then analysed on a 3130xl Genetic Analyzer (Applied Biosystems).

Cloning. GeneArt fragments with mutations were ordered from Invitrogen. Primer sequences can be found in Supplementary Table 4 and GeneArt fragment sequences can be found in Supplementary Table 5. These, along with the miniSmB plasmid (a gift of Dr Benjamin Blencowe), described in ref. 18, were digested with XhoI and NotI and gel purified. Ligation was done using the Quick Ligation Kit (NEB), transformed and selected on ampicillin plates. Minipreps were prepared from selected clones and sequenced at the McGill University and Genome Quebec Innovation Centre.

Transfections. HEK293 cells were plated in 24-well plates at 40–50% confluency. The following day, cells were transfected with 0.25 µg of DNA and 0.75 µl of Fugene 6 (Promega) in 2 ml of DMEM using the manufacturer’s protocol and the resulting purified RNA was resuspended in 20 µl of diethylpyrocarbonate (DEPC)-treated H2O. Transfection was performed three times.

RT-PCR and analysis. Half of the RNA (10 µl) was treated with 2 units of DNase I (NEB) by incubation at 25°C for 15 min, then 1 µl of EDTA 25 mM was added, followed by incubation at 65°C for 10 min. One-eighth of this reaction (2.5 µl) was used in a 10 µl reverse transcriptase reaction using SuperScript III (Invitrogen), according to the manufacturer’s protocol, with 1 µl of cDNA was used in a 10 µl reverse transcriptase reaction using SuperScript III (Invitrogen), according to the manufacturer’s protocol, with 1 µl cDNA, 1 µl of 10 µm primer mix, 10 µl of SYBR Green (Life Technologies) and 8 µl of H2O. Reactions were run on a 7900HT Fast Real-Time PCR System (Applied Biosystems). Cycling conditions were as follows: 95°C for 10 min, 40 cycles of 95°C for 15 s and 60°C for 1 min, and a dissociation step with 95°C for 15 s, 60°C for 15 s and 95°C for 15 s. Relative expression was calculated using the ∆∆CT method36, with EF1β used as a reference gene. The experiment was performed three times. Statistical significance of observed differences was calculated with a Student’s t-test.

qRT-PCR analysis of SNRPB expression. Two microtubes of RNA were used in a 20 µl reverse transcriptase reaction using SuperScript III (Invitrogen), according to the manufacturer’s protocol, with oligo d(T) primers. The resulting cDNA was diluted by one-fifth and used in a qRT-PCR. All qRT-PCRs had a 20 µl volume, with 1 µl cDNA, 1 µl of 10 µm primer mix, 10 µl of SYBR Green (Life Technologies) and 8 µl of H2O. Reactions were run on a 7900HT Fast Real-Time PCR System (Applied Biosystems). Cycling conditions were as follows: 95°C for 10 min, 40 cycles of 95°C for 15 s and 60°C for 1 min, and a dissociation step with 95°C for 15 s, 60°C for 15 s and 95°C for 15 s. Relative expression was calculated using the ∆∆CT method36, with EF1β used as a reference gene. The experiment was performed three times. Statistical significance of observed differences was calculated with a Student’s t-test.

References
1. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint with 29 mammals. Nature 478, 476–482 (2011).
2. Davydov, E. V. et al. Identifying a high fraction of the human genome to be under selective constraint using GERP+++. PLoS Comput. Biol. 6, e1001025 (2010).
3. Bejerano, G. et al. Ultraconserved elements in the human genome. Science 304, 1321–1325 (2004).
4. Dermitzakis, E. T., Reymond, A. & Antonarakis, S. E. Conserved non-genic sequences—an unexpected feature of mammalian genomes. Nat. Rev. Genet. 6, 151–157 (2005).
5. Katzman, S. et al. Human genome ultraconserved elements are ultraselected. Science 317, 915–915 (2007).
6. Encode, T. & Consortium, P. A user’s guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 9, e1001046 (2011).
7. Pennacchio, L. a. et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature 444, 499–502 (2006).
8. Licastro, D. et al. Promiscuity of enhancer, coding and non-coding transcription functions in ultraconserved elements. BMC Genomics 11, 151 (2010).
9. Neph, S. et al. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 489, 83–90 (2012).
10. Sanges, R. et al. Highly conserved elements discovered in vertebrates are present in non-syntonic loci of tunicates, act as enhancers and can be transcribed during development. Nucleic Acids Res. 41, 3600–3618 (2013).
11. Ishibashi, M., Noda, A. O., Sakate, R. & Imanishi, T. Evolutionary growth potential of highly conserved sequences in vertebrate genomes. Gene 504, 1–5 (2012).
12. Pan, Q., Shai, O., Lee, I. I., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. 40, 1413–1418 (2008).
13. Wang, E. T. et al. Alternative isoform regulation in human tissue transcripts. Nature 456, 470–476 (2009).
14. Lareaux, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446, 926–929 (2007).
15. Ni, J. Z. et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. 21, 708–718 (2007).
16. Saltzman, A. L. et al. Regulation of multiple core splicingosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. Mol. Cell. Biol. 28, 4320–4330 (2008).
17. Junna, H. & Nielsen, P. J. The splicing factor SRp20 modifies splicing of its own mRNA and AS/SE/SF2 antagonizes this regulation. EMBO J. 16, 5077–5085 (1997).
18. Saltzman, A. L., Pan, Q. & Blencowe, B. J. Regulation of alternative splicing by the core splicingosomal machinery. Genes Dev. 373–384 (2011).
19. Jones, K. L. Smith’s Recognizable Patterns of Human Malformation 688–689 (Elsevier Ltd, 2006).
20. Kambach, C. et al. Crystal structures of two Sm protein complexes and their implications for the assembly of the splicingosomal snRNPs. Cell 96, 375–387 (1999).
21. Cooper, T. A. Use of minigene systems to dissect alternative splicing elements. Methods 37, 331–340 (2005).
22. Wang, Z. et al. Systematic identification and analysis of exonic splicing silencers. Cell 119, 831–845 (2004).
23. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073 (2010).
ACKNOWLEDGEMENTS

Supporting information can be found in the online Supplementary Material. We thank the patients and their families who participated in this study. Cell culture for the qRT-PCR experiments was performed by Nadine Gamache. The miniSmB construct was a gift from Richard Butz. Extraction and sequencing of the patients and their families who participated in this study. Cell culture for the qRT-PCR experiments was performed by Nadine Gamache. The miniSmB construct was a gift from Richard Butz and the phenotypes associated with EFTUD2 mutations. "Orphanet J. Rare Dis." 8, 110 (2013).

Miller, K. E., Allen, P. & Davis, W. S. Rib gap defects with micrognathia. The cerebro-costo-mandibular syndrome—a pincer brachio-facial syndrome with rib dysplasia. Am. J. Roentgenol. Radial Ther. Nucl. Med. 114, 253–256 (1972).

Gabbanella, F., Carissimi, C., Ussiolo, A. & Pellizzoni, L. The activity of the spinal muscular atrophy protein is regulated during development and cellular differentiation. Hum. Mol. Genet. 14, 3629–3642 (2005).

Tanackovic, G. et al. PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. Hum. Mol. Genet. 20, 2116–2130 (2011).

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).

Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010).

Kym Boycott, Alex MacKenzie, Jacek Majewski, Michael Brudno, Dennis Bulman & David Dyment.

Children’s Hospital of Eastern Ontario Research Institute, University of Ottawa, 401 Smyth Road, Ottawa, Ontario, Canada K1H 8L1. 16Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. 18Donnelly Centre and Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada.

16Children’s Hospital of Eastern Ontario Research Institute, University of Ottawa, 401 Smyth Road, Ottawa, Ontario, Canada K1H 8L1.

17Department of Computer Science, University of Toronto, Toronto, Ontario, Canada.

18Donnelly Centre and Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada.