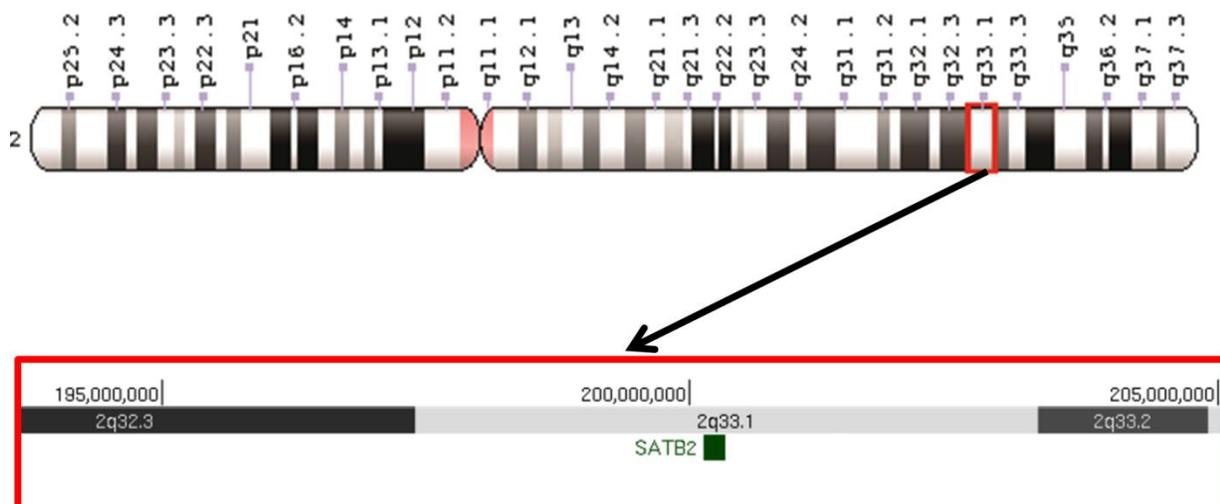


UNDERSTANDING GENETIC TESTING RESULTS

There are different ways changes in our genes might cause *SATB2*-associated syndrome (SAS). These include: contiguous large deletions (pieces missing that include *SATB2* as well as other genes nearby), intragenic deletions (pieces missing of just *SATB2*), large duplications (pieces extra that include *SATB2* as well as other genes), intragenic duplications (pieces extra of just *SATB2*), translocations (chromosome abnormality caused by rearrangement of parts between chromosomes) that include *SATB2*, and mutations (misspellings). We encourage you to read the Glossary section first, so that some of the terms used can be easily understood.

Contiguous large deletions:

Deletions (pieces missing) of several genes are labeled “contiguous”. This is the most commonly reported change involving the *SATB2* gene. Chromosome 2 is the second largest human chromosome spanning about 243 million building blocks of DNA (base pairs). Contiguous large deletions are typically found through a test called a microarray (also known as array CGH or SNP array). The reports are usually written similarly: arr [hg19] 2q33.1 (198,356,789-203,491,035)x1. In this particular example, first we can tell is a deletion by the “x1” at the end that indicates that only one copy is present instead of the normal two (one from mom and one from dad). Then, the “2q33.1” refers to the location where the deletion is. Lastly, the numbers indicate with more precision what exactly is missing, a bit like GPS coordinates. In this case, approximately 5 million letters are missing, abbreviated 5 Mb. One copy of the *SATB2* gene, located in chromosome 2 at position q33.1 around the 200,000,000 position is therefore expected to be missing because the letters located between 198 million and 203 million are missing (See [Figure below](#)). Microarray tests from many years ago may have slightly different numbers that still correspond to the same genes. This is because over the years, as we learn more about DNA, we have to adjust the count of base pairs.



Intragenic deletions:

Deletions (pieces missing) that only include a part of the *SATB2* gene are labeled “intragenic”. Intragenic deletions have only been reported in a handful of patients. An example of an intragenic deletion report from a patient in Dr. Zarate’s study is: arr [hg19] 2q33.1(200,180,939-200,200,560)x1. In this particular example only 2 exons (coding regions of the gene) out of the normal 11, are missing. Because the exons are critical in providing the correct information to form the protein, the *SATB2* gene copy that has that piece missing is unlikely to work properly.

Contiguous large duplications:

Duplications (pieces extra) of multiple genes that include *SATB2* have rarely been reported. It is unclear if having this extra piece results in similar signs and symptoms for patients compared to the other mechanisms listed.

Intragenic duplications:

Duplications (pieces extra) that only include a part of the *SATB2* gene are labeled “intragenic”. Intragenic duplications have only been reported in a handful of patients. An example of an intragenic duplication report from the literature is: arr [hg19] 2q33.1 (200,256,583-200,340,204) x3 (Kaiser et al.). In this particular example the duplication is indicated by the “x3” at the end which means that there are 3 copies of that area instead of the usual two. Only 1 exon (coding regions of the gene) out of the normal 11 is extra in this case. Because the exons are critical in providing the correct information to form the protein, the *SATB2* gene copy that has that extra piece is unlikely to work properly.

Chromosome translocations

Sometimes one chromosome may be attached to another chromosome improperly and be rearranged. This is called a chromosome translocation, and can be a perfectly normal finding or can cause signs and symptoms. Rearrangements between chromosome 2q33.1 and other chromosomes have rarely been reported. If the rearrangement breaks are located within the *SATB2* it could also “break” the gene and make it not work properly.

Mutations inside the gene

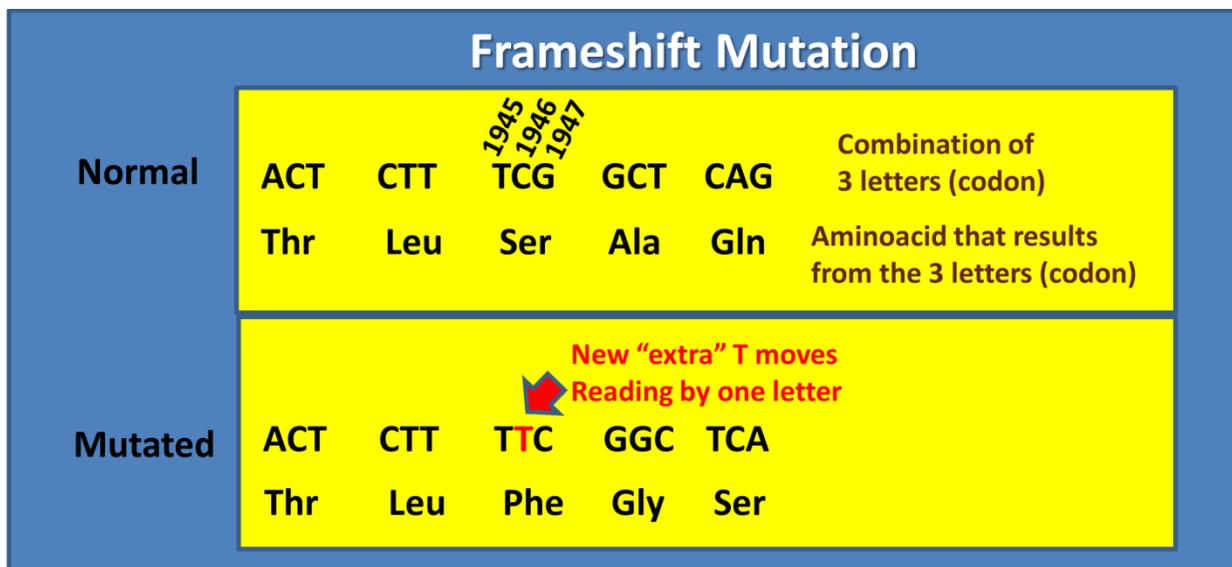
Mutations (Misspellings) of the *SATB2* gene have been reported in a several patients making this the second most frequent cause of SAS. There are different types of mutations, also called pathogenic variants. Some examples from patients enrolled in Dr. Zarate’s study are provided below:

- MISSENSE MUTATIONS: c.346G>C (p.G116R), heterozygous, de novo: Missense variants cause one letter of DNA to be replaced with another letter, and this causes the amino acid to be different at that spot in the protein. In this report the “c.” means coding and the “p” means protein. They represent the same information one describing where the problem is in the gene (the “c”) and the other where the problem is in the protein (the “p”). Counting from the first letter of the gene, the letter 346 is where the problem is. In the usual DNA sequence this letter is supposed to be a G but instead it has been changed to a C. The sequence of 3 letters that is supposed to be read as aminoacid Glycine

(abbreviated G) has now been changed and instead, the new amino acid that results from the new combination of letter is an Arginine (abbreviated R) (See [Figure below](#)). Glycine and Arginine are very different amino acids, and one does not easily fit as a replacement for the other. The number 116 represents the codon (each set of 3 letters) and can be obtained: $3^3=27$, which means that the letter G is the first in the codon 116.

Missense Mutation						
Normal	GCC	CAA	³⁴⁶ ³⁴⁷ ³⁴⁸ GGA	ATA	ATC	Combination of 3 letters (codon)
	Ala	Gln	Gly	Ile	Ile	Amino acid that results from the 3 letters (codon)
Mutated	GCC	CAA	CGA	ATA	ATC	New letter changes the 3 letter combination
	Ala	Gln	Arg	Ile	Ile	

- FRAMESHIFT MUTATIONS:** c.1945dupT (p.S649FfsX40) heterozygous, de novo: Frameshift variants represent a change in the “reading frame”. This means that one or more letters are added or missing from the normal gene and that changes the groups of 3 letters that were supposed to be read. This causes a shift in the sequence that usually makes the protein different and shorter. In this example, there is an extra letter T in the letter number 1945. This extra letter moves all the letters by one position and that results in new combinations of 3 letters resulting in a “shift”. Before the change the amino acid Serine, abbreviated Ser or S, was formed and now the new combination of letters results in a new amino acid named Phenylalanine, abbreviated Phe or F. The “fs” is the abbreviation for frameshift (See [Figure below](#)). Lastly, the report also has a “X40” which means that the protein gets “stopped” 40 amino acids down. As a consequence, the SATB2 protein gets “cut” before it should. So, instead of the 733 amino acids it should have, it now has $649 + 40 = 689$. This is likely to result in a protein that does not work well.



- NONSENSE MUTATIONS:** c.847C>T (p.R283X) heterozygous, de novo: In nonsense variants, a single letter is substituted by another one but the new combination of 3 letters gives the instruction to “cut” the protein on the spot. In this example the letter 847 is where the problem is. In a usual DNA sequence, this letter is supposed to be a C but instead it has been changed to a T. The sequence of 3 letters that is supposed to be read as aminoacid Arginine (abbreviated R) is now replaced by a combination of letters that results in a “stop” codon, abbreviated X or * (See [Figure below](#)). As a consequence, the SATB2 protein gets “cut” before it should and instead of the 733 aminoacids it should have, it now has 283. This is likely to result in a protein that does not work well.

