Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

IDENTIFYING MODIFIER GENES IN SMA MODEL MICE

by
Weiting Xu

Spinal Muscular Atrophy (SMA) involves the loss of nerve cells called *motor neurons* in the spinal cord and is classified as a *motor neuron disease*, it affects 1 in 5000-10000 newborns, one of the leading genetic causes of infant death in USA. Mutations in the *SMN1*, *UBA1*, *DYNC1H1* and *VAPB* genes cause spinal muscular atrophy. Extra copies of the *SMN2* gene modify the severity of spinal muscular atrophy. Mutations in *SMN1* (Motor Neuron 1) mainly causes SMA (Autosomal recessive inheritance). *SMN1* gene mutations lead to a shortage of the SMN protein and SMN protein forms SMN complex which take part in snRNP biogenesis and pre-mRNA splicing. Without SMN protein, motor neurons die, and nerve impulses are not passed between the brain and muscles. As a result, some muscles cannot perform their normal functions, leading to weakness and impaired movement. In this research, we used SMA model mice (LL samples and Sever samples) to identify *de novo* mutations and modifiers operating in SMA model mice.

IDENTIFYING MODIFIER GENES IN SMA MODEL MICE

by
Weiting Xu

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Bioinformatics

Department of Computer Science

May 2015

APPROVAL PAGE

IDENTIFYING MODIFIER GENES IN SMA MODEL MICE

Weiting Xu

Dr. Zhi Wei, Dissertation Advisor Associate Professor of Computer Science, NJIT	Date
Dr. Usman W. Roshan, Committee Member Associate Professor of Computer Science, NJIT	Date
Dr. Antai Wang, Committee Member Associate Professor of Mathematical Sciences, NJIT	Date

BIOGRAPHICAL SKETCH

Author: Weiting Xu

Degree: Master of Science

Date: May 2015

Undergraduate and Graduate Education:

- Master of Science in Bioinformatics,
  New Jersey Institute of Technology, Newark, NJ, 2015

- Bachelor of Science in Biotechnology,
  Northwest University, Xi'an, The People's Republic of China, 2013

Major: Bioinformatics

I dedicate this work to my loved family

ACKNOWLEDGMENT

There are too many words for me to express my gratitude to my master's thesis advisor, Dr. Zhi Wei, for the patient teaching and support on my study and research work in my two years in NJIT. I would like to thank the committee members Dr. Usman W. Roshan and Dr. Antai Wang, for their fabulous teaching and technical help.

I would also like to thank my peers Jie Zhang, Mengnan Gu and Tian Tian, for their help and patient directions.

At the end, I would like to thank my parents, Rui Li and Yongqiang Xu, for their supporting and mental encouragement during my years in NJIT.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Objective and Methods

The objective of this study is to identify Modifier genes in SMA model mice. These tools include FastQC, Bowtie 2, SAMtools, PICARD, GATK and snpEff. Several features such as Mapping Reads, Alignment Manipulation, Variants Calling, Ti/Tv Ratio and Annotation are taken into consideration.

For the Mapping Reads, a table was carried out based on the sequencing reads alignment to reference sequences. Mapping short reads against a reference genome is classically the first step of many next-generation sequencing data analyses, and it should be as accurate as possible[5]. The purpose of mapping is to create an alignment file also known as a Sequence/Alignment Map (SAM) file for each of samples. The SAM file will contain one line for each of the reads in your sample denoting the reference sequence (genes, contigs, or gene regions) to which it maps, the position in the reference sequence, and a Phred-scaled quality score of the mapping. The SAM files can be used for samples to extract gene expression information[1].

For Variants Calling, PICARD tools and GATK are taken into consideration. There are 8 VCF files have been generated to identify sequence variants.

SnpEff is used in Annotation, it is an genetic variant annotation and effect prediction toolbox. It annotates and predicts the effects of variants on genes (such as

amino acid changes). The inputs are predicted variants (SNPs, insertions, deletions and MNPs). The input file is obtained as a result of a sequencing experiment, and it is in variant call format (VCF). SnpEff analyzes the input variants. It annotates the variants and calculates the effects they produce on known genes.

1.2 Background Information

Spinal muscular atrophy (SMA) is a progressive neurodegenerative disorder caused by the loss of function of motor neurons[3]. The loss of motor neurons leads to weakness and wasting (atrophy) of muscles used for activities such as crawling, walking, sitting up, and controlling head movement. In severe cases of spinal muscular atrophy, the muscles used for breathing and swallowing are affected.

Four main types of spinal muscular atrophy affect children before the age of 1. Type I spinal muscular atrophy is a severe form of the disorder that is evident at birth or within the first few months of life. Affected infants are developmentally delayed, most are unable to support their head or sit unassisted and have difficulty breathing and swallowing that may lead to choking or gagging and are unable to sit without support.

Type II spinal muscular atrophy is characterized by muscle weakness that develops in children between ages 6 and 12 months. Children with type II can sit without support, although they cannot stand or walk unaided[1].

Type III spinal muscular atrophy has milder features that typically develop between early childhood and adolescence. Individuals with type III spinal muscular

atrophy can stand and walk unaided, but walking and climbing stairs may become increasingly difficult. Many affected individuals will require wheelchair assistance later in life[7].

The signs and symptoms of type IV spinal muscular atrophy often occur after age 30[8]. Affected individuals usually experience mild to moderate muscle weakness, tremor, twitching, or mild breathing problems. Typically, only muscles close to the center of the body (proximal muscles), such as the upper arms and legs, are affected in type IV spinal muscular atrophy[1][6].

Mutations in the *SMN1*, *UBA1*, *DYNC1H1* and *VAPB* genes cause spinal muscular atrophy. Extra copies of the *SMN2* gene modify the severity of spinal muscular atrophy.

Mutations in the *SMN1* (survival motor neuron 1) gene cause spinal muscular atrophy types I, II, III, and IV. *SMN1* gene mutations lead to a shortage of the SMN protein. Without SMN protein, motor neurons die, and nerve impulses are not passed between the brain and muscles. As a result, some muscles cannot perform their normal functions, leading to weakness and impaired movement.

SMN1 is the primary SMA-related gene. Approximately 95% - 98% of individuals with a clinical diagnosis of SMA are homozygous for a deletion or gene conversion of *SMN1*, typically determined by lack of exon 7 in both copies of *SMN1*. Approximately 2% - 5% of individuals with a clinical diagnosis of SMA are compound heterozygotes for deletion of at least *SMN1* exon 7 and an intragenic inactivating mutation of *SMN1* that is detectable by sequence analysis[4]. Figure 2.1

shows how *SMN1* mutation causes SMA.

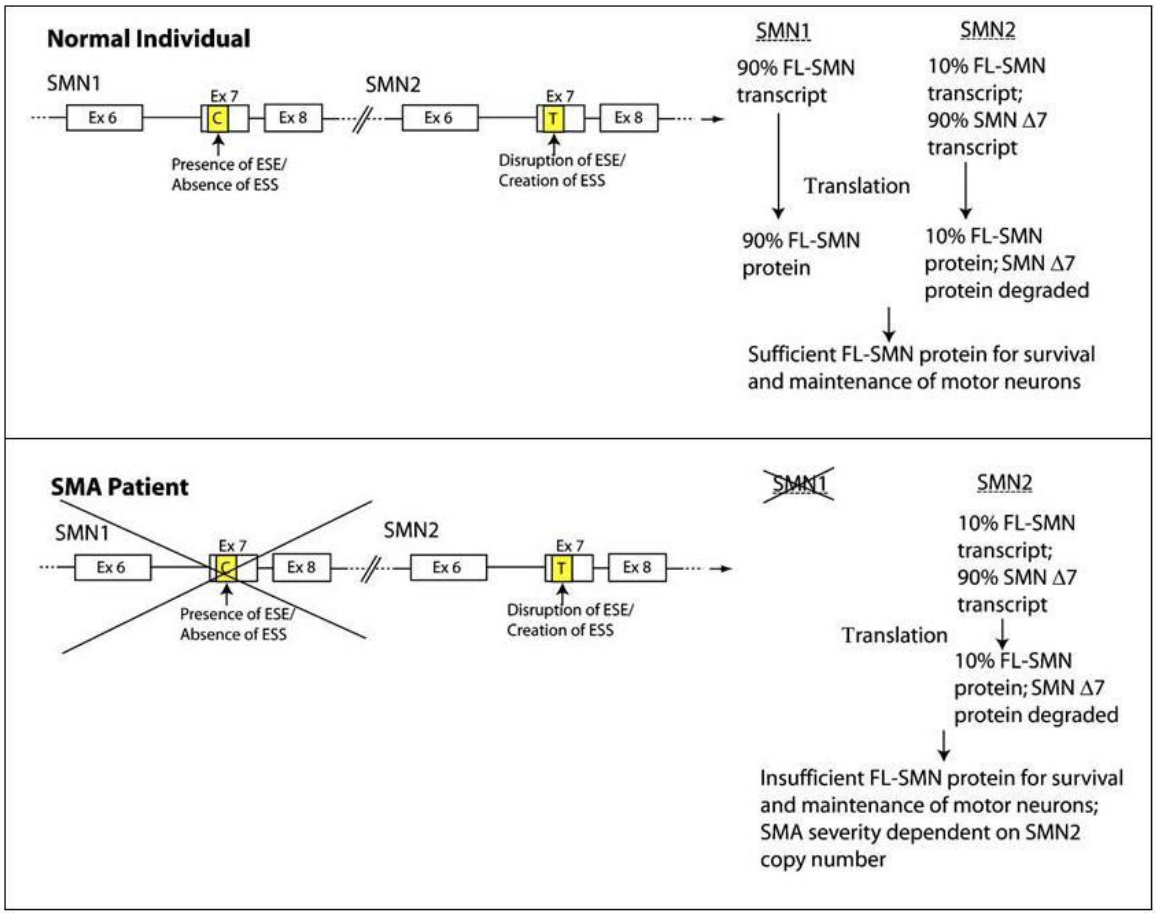

Figure 2.1 The molecular basis of proximal spinal muscular atrophy depicting the two major genes involved in the disease, *SMN1*, and the modifier, *SMN2*.

Source: Monani UR. (2005). Spinal muscular atrophy: a deficiency in a ubiquitous protein; a motor neuron-specific disease. Neuron. 48(6), 885-96.

CHAPTER 2

DATASET

2.1 SMA Model Mice

In this paper experiment, there are eight samples which are divided into two families in SMA model mice. We used LL samples (Long Live samples) and sever samples which came from same parents pair for exome sequencing. Our LL and typical SMA mice are a mix of the FVB/N and C57BL/6 strains of mice. FVB/N mice offer a system suitable for most transgenic experiments and subsequent genetic analyses[10]. C57BL/6 is a common inbred strain of laboratory mouse.It is the most widely used "genetic background" for genetically modified mice for use as models of human disease. Figure 2.2 shows our experiment for SMA Model Mice vs Control Mice within different time range.

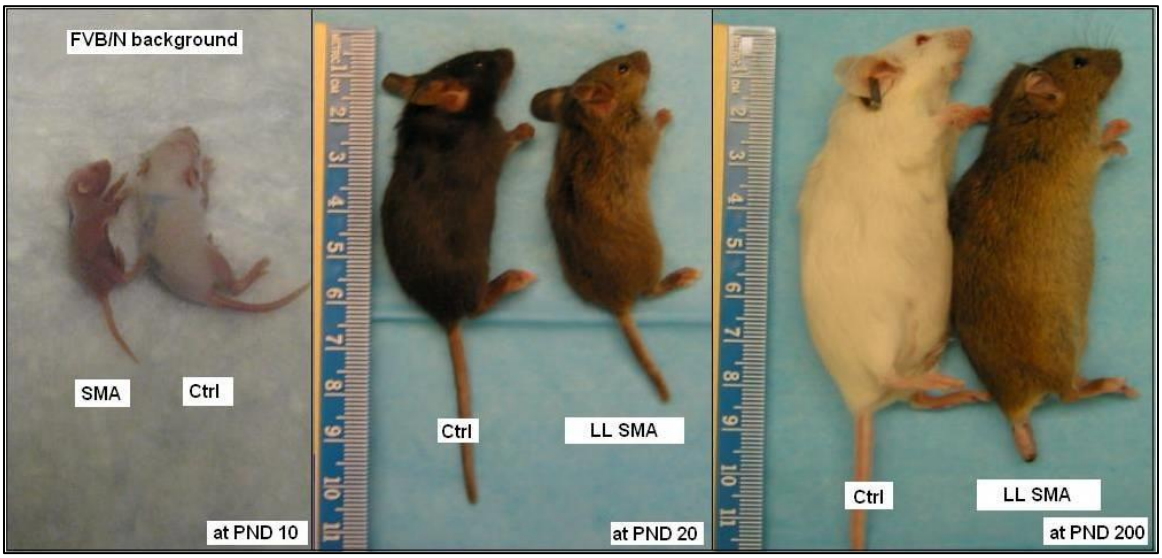

Figure 2.2 Appearance results within different time range.

Source: From our cooperation Columbia University Biology Lab

Figure 2.3 shows Kaplan-Meier survival curve analysis and it indicates a highly significant effect of modifiers in an F1 intercross between C57Bl/6 and FVB/N compared to factors in FVB/N alone (2=7.4; P<0.005).

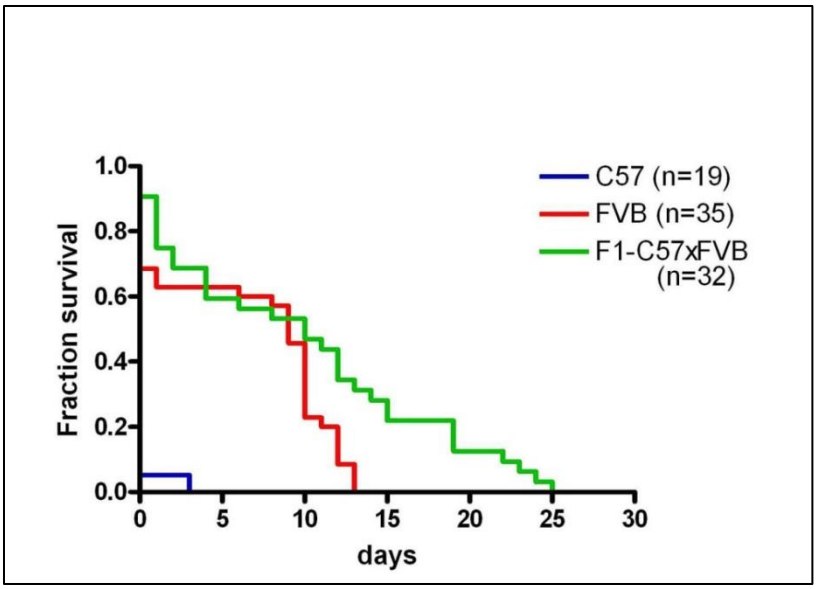

Figure 2.3 Kaplan-Meier survival curve analysis.

Source: From our cooperation Columbia University Biology Lab

Table 2.1 shows the probability of LL mutants in each generation. The detailed information about data in this experiment is showed in Table 2.2. Family #1: UM01, UM02 and UM03 are LL samples, UM05 and UM07 are sever samples. Family #2: UM11 and UM12 are LL samples, and UM13 is sever sample.

Table 2.1 Probability of LL Mutants in Each Generation

Generation	Probability of LL Mutants Per All Mutants	Number of Samples
F1xF1	4%	1 out of 25
F2xF2	13.3%	6 out of 45
F3xF3	27.9%	19 out of 68

If modifiers works as autosomal recessive manner, 1 modifier become homozygous 25%, 2 modifiers become homozygous 6.25%, 3 modifiers become

homozygous 1.56%.

1. In F1xF1 generation, the probability of LL is 4%. Therefore, we estimated that at least 2 modifiers (most likely 3 modifiers) are existing is LL mutants.

2. In F2xF2 and F3xF3 generations, probability of LL was increased due to the biased selection of breeding pairs (the breeding pair which produce LL mice were more frequently used for breeding.)

3. In F3xF3 generation, probability of LL is above 25%. This is indicating that 2 modifiers (if 3 mods exist) might be already homozygously fixed in breeding pair.

Table 2.2 Description of the SMA Model Mice

	LL Samples	**Sever Samples**
Family #1	UM01, UM02, UM03	UM05, UM07
Family #2	UM11, UM12	UM13

CHAPTER 3

ALIGNMENT AND COUNTS

3.1 Mapping Reads

Bowtie 2 is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences. It is particularly good at aligning reads of about 50 up to 100s or 1,000s of characters, and particularly good at aligning to relatively long genomes. Bowtie 2 is often the first step in pipelines for comparative genomics, including for variation calling, ChIP-seq, RNA-seq, BS-seq. Multiple processors can be used simultaneously to achieve greater alignment speed.

We use Bowtie 2 for aligning our mouse data to mm9 fasta reference data and select reads that have a mapping quality of 20 or better (DP > 20) and then count the amount of those mapped reads and save it in the Table 3.1 Amount of mapped reads. Next step we use SAMtools to count the amount of mapped reads in target region which named S0276129_Regions bed file download from SureDesign website and we save the count number in the Table 3.1 Amount of mapped reads in target region. SAMtools provide various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format.

Table 3.1 Summary of the Mapping Reads

ID	Amount of mapped reads	Amount of mapped reads in target region
UM01	48055975	36786262
UM02	45410389	30674883
UM03	59372714	47783279
UM05	42657431	34684432
UM07	40806247	30727583
UM11	41675811	31039464
UM12	62132734	48285532
UM13	40781358	31228352

3.2 Evaluation by Ti/Tv Ratio

Ti/Tv ratio is also known as Transition-Transversion ratio. This is a ratio of the number of transition to transversion substitutions that appear to have occurred since two sequences separated from a common ancestor. It is also the average rate of transition versus transversion substitutions in a dataset. What's more, Ti/Tv ratio is a value, estimated by reference to a tree that describes the average rate of transition to transversion substitutions during the evolutionary period covered by the tree. Estimation of the ti/tv rate bias is important not only to our understanding of the patterns of DNA sequence evolution, but also to reliable estimation of sequence distance and phylogeny reconstruction[9].

The Figure 3.2 is an illustration. AG, CT pairs are defined as transitions, while other pairs are considered as transversions. The Table 3.2 we calculate variants Ti/Tv Ratio in target region (S0276129_Regions bed file). Expected human Ti/Tv ratio for

whole-genome is 2.1 and for whole-exome is 3.0, FP SNPs should have a Ti/Tv of 0.5. Note that these expectations are only for individual sequencing data. It may not be held for pooled sequencing data. There is no expected Ti/Tv ratio for mouse now so we just calculate it for a reference.

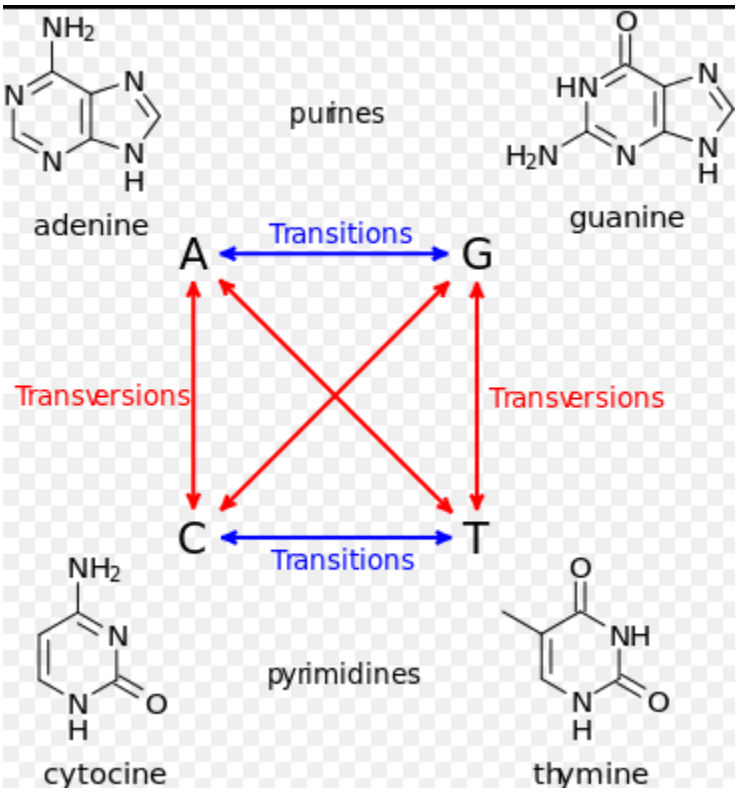

Figure 3.2 Definition of transitions and transversions.

Source: http://en.wikipedia.org/wiki/Transversion Transversion on Wikipedia

Table 3.2 Ti/Tv Ratio

ID	Ti/Tv Ratio in target region
UM01	2.879739
UM02	1.754267
UM03	2.790435
UM05	2.889236
UM07	2.750107
UM11	2.224619
UM12	2.329107
UM13	2.121361

CHAPTER 4

RESULTS

4.1 Alignment Manipulation and Variants Calling

After we finish the alignment mapping, the next step is to do the alignment manipulation and variants calling. We utilize several PICARD tools to summarize the alignments and use GATK to analyze data. We have 8 VCF files which contains specific variants after these steps. The Table 4.1 shows the description of each column in VCF file.

Picard comprises Java-based command-line utilities that manipulate SAM files, and a Java API (HTSJDK) for creating new programs that read and write SAM files. Both SAM text format and SAM binary (BAM) format are supported. We choose several PICARD tools:

1. CollectAlignmentSummaryMetrics, it reads a SAM or BAM file and writes a file containing summary alignment metrics.

2. MarkDuplicates, it examines aligned records in the supplied SAM or BAM file to locate duplicate molecules. All records are then written to the output file with the duplicate records flagged.

3. AddOrReplaceReadGroups, it replaces all read groups in the INPUT file with a new read group and assigns all reads to this read group in the OUTPUT. Note that this step is required by the latest GATK, which no longer supports SAM files without read groups.

4. BuildBamIndex, generates a BAM index (.bai) file using PICARD.

After the PICARD we do the Variants Detection by GATK. The Genome Analysis Toolkit (GATK) is a software package developed at the Broad Institute to analyze

next-generation sequencing data. The toolkit offers a wide variety of tools, with a primary focus on variant discovery and genotyping as well as strong emphasis on data quality assurance. It consists of the following steps:

1. Call variants by UnifiedGenotyper, we use the GATK Unified Genotyper to detect variants, which is based on a Bayesian genotype likelihood model. The variant calls are stored in VCF (Variant Call Format).

2. Select variants from a VCF source, a VCF containing many samples and/or variants will need to be subset in order to facilitate certain analyses (e.g. comparing and contrasting cases vs. controls). SelectVariants can be used for this purpose. Given a single VCF file, one or more samples can be extracted from the file (based on a complete sample name or a pattern match). Variants can be further selected by specifying criteria for inclusion. In this experiment we use SelectVariants to select variants which has “DP > 20” (depth of coverage greater than 20x).

3. We use intersect function to select variants in target region (S0276129_Regions bed file). The variant calls are stored in VCF.

Table 4.1 Description of Each Column in Output VCF File

VCF Column	Description
1st	chromosome number
2nd	snp location
3rd	ID: unique identifier
4th	reference base
5th	sample base
6th	phred-scaled quality score
7th	filter
8th	Additional information

4.2 Annotation

We use those variants that have DP>20 (depth of coverage greater than 20x) and intersected reads in target region (S0276129_Regions bed file) to do annotate. We chose snpEff tool.

SnpEff is a variant annotation and effect prediction tool. It annotates and predicts the effects of variants on genes (such as amino acid changes). The inputs are predicted variants (SNPs, insertions, deletions and MNPs). The input file is usually obtained as a result of a sequencing experiment, and it is usually in variant call format (VCF). SnpEff analyzes the input variants. It annotates the variants and calculates the effects they produce on known genes (e.g. amino acid changes). SnpEff updates the header of the VCF file to reflect additional fields. It also adds the command line options used to annotate the file as well as SnpEff's version, so we can keep track of what exactly was done.

4.3 Filtering

We use previous variants VCF file to do filtering to find those variants that will impact protein functions and follow inheritance model (recessive model). The most recent analysis which our cooperative team members carried out using a panel of 1500 SNPs at the Jackson Labs suggests that there is a major modifier of the SMA phenotype on chromosome 9 of the mouse genome. Accordingly, we have been scouring the genes in

this region for variants between FVB/N and C57Bl6 for further clues. One caveat of this strategy is that the modifier arise *de novo* in our colony of mice. If this is the case, we will miss it by simply comparing sequences in the publicly available databases. In essence then, the idea would be to compare sequence variants (initially only protein coding) on chromosome 9 in our long-lived mice using C57Bl6 as the reference sequence on the one hand and FVB/N as the reference sequence on the other. Any *de novo* alterations in our mice would then become apparent and can be combined with our SNP analysis data to further home in on potential individual modifiers. Then, the idea consists of the following steps:

1. Identify chromosome 9 variants in LL samples, i.e., UM01, UM02, UM03, UM11, UM12. Limit these to variants that either alter the coding sequence of the protein or the splice sites of the gene. So we select chr 9 first from variants get from Annotation in LL samples and then we select missense_variant and splice_region_variant.

2. Determine which of the above variants are common to all LL samples.

3. Identify chromosome 9 variants in all typical samples, i.e., UM05, UM07, UM13. Limit these to variants that either alter the coding sequence of the protein or the splice sites of the gene. So we select chr 9 first from variants get from Annotation in typical samples and then we select missense_variant and splice_region_variant.

4. Determine which of the variants identified in step 3 are common to all typical samples.

5. Identify which of the variants in step 3 are present only in the typical samples and not in the LL samples. Further identify which of the variants from step 1 are present in all or a majority of the LL samples but absent in the typical samples. Here we use R packages to do the intersection and set difference. The Table 4.3 shows variants present only in the typical samples and not in the LL samples; The Table 4.4 shows variants present in all LL samples but absent in the typical samples.

6. Using variants identified in step 1, determine which of the variants in the LL mice derive from the FVB/N strain of mice.

7. Remove variants identified in step 1 that derive from the FVB/N strain. This will leave variants that are most likely *de novo* mutations, i.e., those that arose in in our colony of mice. The Table 4.5 shows variants that are most likely *de novo* mutations.

8. Determine which of the possible *de novo* mutations identified in step 7 are common to all LL samples. Identify which, if any, of these are also present in the typical samples. The Table 4.6 shows possible *de novo* mutations identified in step 7 are common to all LL samples and are also present in the typical samples.

CHAPTER 5

CONCLUSION

Our LL and typical SMA mice are a mix of the FVB/N and C57Bl/6 strains of mice. Based on our mapping studies and on the incidence of the LL mice, we expect a total of 2-3 modifiers at least one of which is recessively inherited.

As expected, from Table 4.3 we found 55 variants, 18 variants in UM05, 5 variants in UM07 and 32 variants in UM13. From Table 4.4 we found 335 variants, 5 variants in UM01, 224 variants in UM02, 7 variants in UM03, 5 variants in UM11 and 94 variants in UM12. From Table 4.5 we found 506 variants, 70 variants in UM01, 243 variants in UM02, 20 variants in UM03, 62 variants in UM11 and 111 variants in UM12. And from Table 4.6 we only found 2 variants left, these two are possible *de novo* mutations and modifiers operating in SMA model mice.

APPENDIX A

POSSIBLE *de novo* MUTATIONS AND MODIFIERS OPERATING IN SMA MODEL MICE

Possible *de novo* mutations and modifiers operating in SMA model mice are provided in the following Tables.

Table A.1 UM07 Variants Present Only in the Typical Samples and not in the LL Samples

		UM07		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	49213127	Drd2	C	A
chr9	55062707	Fbxo22	C	G
chr9	78179006	Gsta2	A	C
chr9	87134697	Cep162	G	A
chr9	103917649	Nphp3	T	A

Table A.2 UM05 Variants Present Only in the Typical Samples and not in the LL Samples

		UM05		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	38238446	Olfr901	C	G
chr9	56108186	Peak1	C	A
chr9	56108188	Peak1	C	T
chr9	86489810	Me1	A	G
chr9	92164821	Plscr1	G	C
chr9	95582036	Pcolce2	G	A
chr9	95593325	Pcolce2	T	C
chr9	95765989	Atr	C	G
chr9	96233950	Atp1b3	C	T
chr9	96234031	Atp1b3	T	C
chr9	96234041	Atp1b3	C	T
chr9	96587156	Zbtb38	G	A
chr9	96587405	Zbtb38	C	T
chr9	97355119	Clstn2	A	C
chr9	97355144	Clstn2	T	G
chr9	99231120	Esyt3	C	T
chr9	100386717	Il20rb	A	T
chr9	101831601	Ephb1	C	T

Table A.3 UM13 Variants Present Only in the Typical Samples and not in the LL Samples

		UM13		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	19180243	Olfr847	G	C
chr9	19180247	Olfr847	A	T
chr9	19180248	Olfr847	G	T
chr9	20065163	Olfr872	C	G
chr9	23203898	Bmper	T	A
chr9	37892652	Olfr887	T	G
chr9	37892653	Olfr887	A	G
chr9	37892654	Olfr887	A	T
chr9	37892655	Olfr887	C	T
chr9	39647108	Olfr971	C	T
chr9	39647114	Olfr971	C	T
chr9	39681067	Olfr972	T	A
chr9	39681072	Olfr972	C	G
chr9	39681073	Olfr972	T	G
chr9	44317448	Bcl9l	A	T
chr9	45510729	Dscaml1	G	T
chr9	45510731	Dscaml1	A	T
chr9	48823237	Usp28	G	T
chr9	54863461	Chrna3	G	T
chr9	55285397	AI118078	C	T
chr9	59513732	Pkm	C	T
chr9	59513733	Pkm	G	C
chr9	65642524	Zfp609	C	T
chr9	89862738	Rasgrf1	C	A
chr9	92161439	Plscr1	G	A
chr9	95609223	Trpc1	A	T
chr9	100397792	Nck1	C	A
chr9	103183362	1300017J02Rik	C	A
chr9	103183365	1300017J02Rik	C	A
chr9	106760944	Vprbp	C	A
chr9	108009008	Bsn	C	T

chr9	108012524	Bsn	A	C

Table A.4 UM01 Variants Present in all LL Samples but Absent in the Typical Samples

		UM01		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	32064570	Arhgap32	C	G
chr9	79991493	Senp6	A	G
chr9	87141166	Cep162	T	C
chr9	90120887	Tbc1d2b	T	C
chr9	122834341	Zfp105	A	C

Table A.5 UM03 Variants Present in all LL Samples but Absent in the Typical Samples

		UM03		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	20867155	Zglp1	C	A
chr9	56108190	Peak1	C	T
chr9	61804006	Paqr5	C	A
chr9	75195821	Bcl2l10	G	A
chr9	75238180	Mapk6	G	T
chr9	75238181	Mapk6	G	C
chr9	1.23E+08	Zfp105	A	C

Table A.6 UM02 Variants Present in all LL Samples but Absent in the Typical Samples

		UM02		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	4330888	Kbtbd3	G	A
chr9	5321472	Casp4	G	T
chr9	5321474	Casp4	C	T
chr9	6265244	Ddi1	C	A
chr9	6265245	Ddi1	A	C
chr9	7005535	Dync2h1	C	G
chr9	7005537	Dync2h1	A	T
chr9	7142292	Dync2h1	C	A
chr9	7451731	Mmp3	T	A
chr9	7854589	Birc3	C	T
chr9	8099848	AK129341	T	A
chr9	8652980	Trpc6	G	A
chr9	8652983	Trpc6	C	A
chr9	9673889	Cntn5	C	T
chr9	14604060	Mre11a	G	T
chr9	15122153	Cep295	T	A
chr9	15127320	Cep295	G	A
chr9	15136690	Cep295	G	T
chr9	15136691	Cep295	G	A
chr9	15136693	Cep295	T	A
chr9	15136694	Cep295	G	A
chr9	15140072	Cep295	T	A
chr9	15140073	Cep295	C	G
chr9	15140075	Cep295	T	G
chr9	15719744	Fat3	G	T
chr9	15719745	Fat3	G	A
chr9	15808505	Fat3	G	A
chr9	18105499	Chordc1	G	A
chr9	18183070	Naalad2	C	A
chr9	18249145	Mbd3l2	G	T
chr9	18289497	Mbd3l1	C	T
chr9	18620250	Olfr828	C	A

chr9	18749821	Olfr832	C	T
chr9	19371266	Olfr854	C	G
chr9	19612851	Olfr859	C	A
chr9	19612854	Olfr859	T	A
chr9	19612858	Olfr859	T	A
chr9	19724728	Olfr77	C	A
chr9	19725116	Olfr77	G	T
chr9	19905202	Olfr868	G	A
chr9	20065081	Olfr872	T	A
chr9	20576761	Col5a3	G	A
chr9	20813042	Mrpl4	C	G
chr9	20813043	Mrpl4	C	T
chr9	20830828	Icam1	G	T
chr9	24279840	Dpy19l1	G	T
chr9	31137256	Prdm10	G	T
chr9	31217945	Nfrkb	T	A
chr9	31217947	Nfrkb	C	T
chr9	32056163	Arhgap32	A	T
chr9	32064402	Arhgap32	G	A
chr9	32203916	Kcnj1	T	A
chr9	35019942	Srpr	C	A
chr9	37044551	Slc37a2	C	G
chr9	37833933	Olfr883	C	A
chr9	38017436	Olfr893	C	T
chr9	38017438	Olfr893	G	A
chr9	38137762	Olfr25	C	T
chr9	38256498	Olfr902	C	A
chr9	38256501	Olfr902	G	A
chr9	38256966	Olfr902	G	T
chr9	38257198	Olfr902	G	A
chr9	38324354	Olfr908	T	A
chr9	38401852	Olfr913	G	C
chr9	38401853	Olfr913	G	A
chr9	38402124	Olfr913	T	A
chr9	38402127	Olfr913	T	A
chr9	38480786	Olfr918	A	T

chr9	38505759	Olfr919	C	A
chr9	38505932	Olfr919	G	T
chr9	38685070	Olfr926	T	A
chr9	38867838	Olfr937	A	T
chr9	38951907	Olfr27	C	A
chr9	39025727	Olfr944	C	A
chr9	39202209	Olfr951	T	A
chr9	39202210	Olfr951	C	A
chr9	39202212	Olfr951	G	A
chr9	39358175	Olfr958	A	T
chr9	39544463	Olfr150	C	T
chr9	39544592	Olfr150	G	T
chr9	39544593	Olfr150	C	G
chr9	39544595	Olfr150	G	A
chr9	39558019	Olfr967	G	A
chr9	39558021	Olfr967	T	A
chr9	39558031	Olfr967	C	A
chr9	39681268	Olfr972	T	A
chr9	39681269	Olfr972	G	T
chr9	40613062	Hspa8	A	C
chr9	43854739	Thy1	C	T
chr9	44432691	Ddx6	G	T
chr9	44629233	Kmt2a	C	A
chr9	45680035	Rnf214	C	T
chr9	46100258	Bud13	G	A
chr9	48128691	Nxpe2	C	A
chr9	48128744	Nxpe2	T	C
chr9	48134429	Nxpe2	C	G
chr9	48134430	Nxpe2	A	G
chr9	49086802	Gm4894	C	A
chr9	49086804	Gm4894	T	A
chr9	49210350	Drd2	C	T
chr9	49210353	Drd2	C	T
chr9	50608230	Alg9	A	T
chr9	50608232	Alg9	C	A
chr9	51647028	Arhgap20	G	A

chr9	53308901	Atm	C	A
chr9	53335437	Atm	G	A
chr9	53371037	Npat	T	A
chr9	54264543	Dmxl2	T	A
chr9	54275846	Dmxl2	C	A
chr9	56105408	Peak1	C	A
chr9	56107575	Peak1	C	A
chr9	56107577	Peak1	C	A
chr9	56107579	Peak1	T	G
chr9	56107580	Peak1	T	A
chr9	56740315	Cspg4	A	T
chr9	57101972	1700017B05Rik	C	T
chr9	58488505	Nptn	A	C
chr9	59627997	Myo9a	C	T
chr9	59628006	Myo9a	A	T
chr9	59680060	Myo9a	C	A
chr9	63372766	Iqch	T	A
chr9	64982405	Igdcc4	G	T
chr9	65323649	Spg21	T	A
chr9	66102402	Dapk2	A	G
chr9	66281948	Herc1	C	A
chr9	66318169	Herc1	G	T
chr9	66803713	Lactb	T	A
chr9	69263617	Ice2	C	T
chr9	69263620	Ice2	T	A
chr9	69844903	Bnip2	A	T
chr9	70277410	Rnf111	T	A
chr9	70277411	Rnf111	C	G
chr9	70291401	Rnf111	G	A
chr9	70434520	Sltm	G	T
chr9	72178796	Zfp280d	C	G
chr9	72178798	Zfp280d	C	A
chr9	72464565	Rfx7	A	T
chr9	72465438	Rfx7	C	T
chr9	72587340	Nedd4	G	T
chr9	72587341	Nedd4	T	A

chr9	72793276	Pygo1	G	C
chr9	72886431	Pigb	T	A
chr9	73780131	Unc13c	T	A
chr9	74066394	Wdr72	C	T
chr9	74737323	Onecut1	G	T
chr9	74737325	Onecut1	A	T
chr9	77639876	Gclc	G	A
chr9	77639878	Gclc	C	A
chr9	77823610	Elovl5	G	A
chr9	78308647	Mto1	T	G
chr9	78308649	Mto1	G	A
chr9	78508747	Cd109	G	C
chr9	78528716	Cd109	G	A
chr9	79541172	Col12a1	T	A
chr9	79551638	Col12a1	C	G
chr9	79551641	Col12a1	C	T
chr9	80129398	Myo6	G	A
chr9	80193399	Impg1	T	A
chr9	85604075	Ibtk	G	A
chr9	85622103	Ibtk	A	T
chr9	85622104	Ibtk	C	A
chr9	85622105	Ibtk	C	A
chr9	85637159	Ibtk	C	A
chr9	85637299	Ibtk	C	T
chr9	86406201	Dopey1	G	C
chr9	87088515	Cep162	C	T
chr9	87143279	Cep162	C	G
chr9	87143317	Cep162	A	T
chr9	87143320	Cep162	G	T
chr9	88293202	Snx14	G	T
chr9	88293204	Snx14	G	C
chr9	88371744	Syncrip	C	A
chr9	88371746	Syncrip	C	G
chr9	89491399	AF529169	G	T
chr9	89497377	AF529169	C	A
chr9	89839643	Rasgrf1	A	C

chr9	89839644	Rasgrf1	C	A
chr9	92185938	Plscr2	T	G
chr9	92490170	Plod2	T	C
chr9	95396352	U2surp	G	T
chr9	95878281	Xrn1	A	T
chr9	95882343	Xrn1	G	T
chr9	95939143	Xrn1	C	T
chr9	95939144	Xrn1	A	G
chr9	96240725	Atp1b3	C	T
chr9	96588607	Zbtb38	G	A
chr9	98481612	Copb2	T	A
chr9	98481614	Copb2	C	A
chr9	98804165	7420426K07Rik	C	A
chr9	98973990	Pik3cb	T	C
chr9	99479865	Dbr1	T	C
chr9	99479868	Dbr1	G	T
chr9	99617481	Cldn18	A	T
chr9	100781294	Stag1	C	A
chr9	101113913	Ppp2r3a	C	A
chr9	101831624	Ephb1	A	T
chr9	103230742	Topbp1	G	C
chr9	103907843	Nphp3	T	A
chr9	105842055	Col6a5	C	A
chr9	106197534	Poc1a	T	A
chr9	106387469	Rrp9	G	A
chr9	106760608	Vprbp	C	A
chr9	106760610	Vprbp	G	A
chr9	106881075	Dock3	G	T
chr9	107667818	Rbm5	T	A
chr9	108307741	Ccdc36	T	A
chr9	108465610	Impdh2	G	A
chr9	108482164	P4htm	T	C
chr9	108698628	Ip6k2	T	A
chr9	108698629	Ip6k2	T	A
chr9	108872553	Col7a1	G	T
chr9	108992003	Ccdc51	G	A

chr9	109794260	Cdc25a	T	A
chr9	110420128	Kif9	C	A
chr9	110451909	Setd2	C	A
chr9	110495073	Setd2	C	T
chr9	111293250	Trank1	G	T
chr9	111474884	Stac	C	T
chr9	115155165	Stt3b	T	A
chr9	122834341	Zfp105	A	C
chr9	122857133	1110059G10Rik	G	A
chr9	122857136	1110059G10Rik	C	A
chr9	123479114	Sacm1l	G	T
chr9	123943822	Ccr3	C	A

Table A.7 UM11 Variants Present in all LL Samples but Absent in the Typical Samples

		UM11		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	32064570	Arhgap32	C	G
chr9	79991493	Senp6	A	G
chr9	87141166	Cep162	T	C
chr9	90120887	Tbc1d2b	T	C
chr9	1.23E+08	Zfp105	A	C

Table A.8 UM12 Variants Present in all LL Samples but Absent in the Typical Samples

		UM12		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	3458760	Cwf19l2	G	A
chr9	4472124	Gria4	G	A
chr9	4472125	Gria4	C	A
chr9	6265403	Ddi1	C	A
chr9	7447622	Mmp3	T	A
chr9	7447625	Mmp3	A	C
chr9	14888710	Hephl1	A	G
chr9	18679785	Olfr830	G	C
chr9	18680357	Olfr830	C	A
chr9	18680364	Olfr830	A	T
chr9	18839764	Olfr835	T	A
chr9	19905445	Olfr868	G	A
chr9	20090340	Olfr39	C	T
chr9	37663042	Olfr877	C	A
chr9	37705250	Olfr145	A	T
chr9	37705253	Olfr145	A	G
chr9	37705255	Olfr145	A	T
chr9	37727015	Olfr878	T	A
chr9	37892427	Olfr887	G	A
chr9	37892918	Olfr887	A	T
chr9	37892919	Olfr887	C	T
chr9	37916766	Olfr888	G	C
chr9	37916769	Olfr888	G	A
chr9	37916770	Olfr888	C	A
chr9	37916773	Olfr888	G	T
chr9	37916774	Olfr888	C	G
chr9	38017396	Olfr893	T	A
chr9	38061670	Olfr143	A	G
chr9	38061709	Olfr143	G	A
chr9	38137186	Olfr25	G	A
chr9	38137187	Olfr25	C	T
chr9	38156679	Olfr898	A	C

chr9	38210776	Olfr147	C	A
chr9	38210782	Olfr147	C	A
chr9	38271775	Olfr904	G	A
chr9	38402127	Olfr913	T	A
chr9	38623387	Olfr922	C	A
chr9	38655923	Olfr924	A	G
chr9	38655924	Olfr924	C	A
chr9	38802432	Olfr935	T	G
chr9	38867955	Olfr937	C	G
chr9	38867956	Olfr937	C	A
chr9	38867982	Olfr937	T	C
chr9	38991986	Olfr943	C	A
chr9	39065420	Olfr945	G	A
chr9	39269838	Olfr954	G	A
chr9	39318982	Olfr957	C	T
chr9	39319039	Olfr957	T	G
chr9	39319044	Olfr957	A	T
chr9	39431016	Olfr960	T	G
chr9	39454395	Olfr961	T	A
chr9	39454397	Olfr961	T	A
chr9	39454398	Olfr961	T	G
chr9	39454399	Olfr961	T	A
chr9	39454400	Olfr961	T	C
chr9	39603161	Olfr969	C	A
chr9	39627436	Olfr970	C	A
chr9	39627956	Olfr970	A	C
chr9	39628040	Olfr970	A	T
chr9	40610828	Hspa8	A	T
chr9	40612763	Hspa8	G	T
chr9	45719163	Pcsk7	T	G
chr9	54264543	Dmxl2	T	A
chr9	54731831	Ireb2	T	C
chr9	54754461	Ireb2	C	A
chr9	55062811	Fbxo22	A	T
chr9	56818116	Snupn	T	C
chr9	57872976	Cyp11a1	G	A

chr9	58725960	Neo1	G	T
chr9	66349065	Herc1	C	G
chr9	70258022	Ccnb2	G	A
chr9	72462964	Rfx7	T	A
chr9	72462966	Rfx7	C	A
chr9	73546927	Unc13c	C	A
chr9	75241169	Mapk6	T	G
chr9	78176355	Omt2b	T	C
chr9	79528174	Col12a1	C	A
chr9	79528175	Col12a1	C	T
chr9	80110160	Myo6	G	T
chr9	85604075	Ibtk	G	A
chr9	85604076	Ibtk	C	A
chr9	85604077	Ibtk	C	A
chr9	85604078	Ibtk	T	A
chr9	85637086	Ibtk	C	T
chr9	87122023	Cep162	G	T
chr9	95364782	U2surp	C	T
chr9	1.03E+08	Topbp1	G	T
chr9	1.03E+08	Topbp1	G	T
chr9	1.04E+08	Dnajc13	C	T
chr9	1.06E+08	Col6a4	C	T
chr9	1.1E+08	Setd2	A	C
chr9	1.14E+08	Ubp1	A	T
chr9	1.19E+08	Itga9	C	A
chr9	1.24E+08	Ccr5	G	A

Table B.9 UM01 Variants Are Most Likely *de novo* Mutations

		UM01		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	32256985	Arhgap32	C	G
chr9	55209076	Fbxo22	G	A
chr9	55209101	Fbxo22	G	T
chr9	55209382	Fbxo22	G	A
chr9	55221070	Fbxo22	A	C
chr9	55221151	Fbxo22	G	T
chr9	75388680	Mapk6	T	A
chr9	78053997	Gcm1	C	T
chr9	78377563	Ooep	C	T
chr9	79647575	Col12a1	A	G
chr9	80093551	Senp6	G	C
chr9	80116600	Senp6	A	C
chr9	80130844	Senp6	G	C
chr9	80143686	Senp6	A	G
chr9	80316176	Impg1	G	T
chr9	80394203	Impg1	T	A
chr9	80465254	Impg1	C	A
chr9	80465281	Impg1	C	A
chr9	85719091	Ibtk	A	G
chr9	85732671	Ibtk	C	A
chr9	85844637	Tpbg	C	T
chr9	85844843	Tpbg	C	A
chr9	86502989	Dopey1	G	A
chr9	86562676	Pgm3	C	G
chr9	86586988	Me1	C	T
chr9	86815426	Snap91	A	G
chr9	87040451	Cyb5r4	A	G
chr9	87057232	Cyb5r4	G	A
chr9	87058957	Cyb5r4	G	A
chr9	87217153	Cep162	T	C
chr9	87220446	Cep162	T	C
chr9	87225720	Cep162	C	T

chr9	87225946	Cep162	T	G
chr9	87227270	Cep162	T	C
chr9	87231408	Cep162	C	T
chr9	87246331	Cep162	T	C
chr9	87248431	Cep162	A	G
chr9	88364661	Nt5e	T	C
chr9	89915497	Rasgrf1	G	A
chr9	90064250	Ctsh	A	T
chr9	90193461	Adamts7	G	T
chr9	90193821	Adamts7	C	T
chr9	90226049	Tbc1d2b	T	C
chr9	90227473	Tbc1d2b	T	C
chr9	104018241	Nphp3	A	G
chr9	104024628	Nphp3	A	C
chr9	104024641	Nphp3	C	T
chr9	104024658	Nphp3	A	G
chr9	104029997	Nphp3	A	G
chr9	104031987	Nphp3	G	A
chr9	104033409	Nphp3	G	A
chr9	104075894	Acad11	A	C
chr9	104076435	Acad11	A	G
chr9	118572989	Golga4	A	C
chr9	122857512	Zfp445	A	G
chr9	122861850	Zfp445	C	G
chr9	122861902	Zfp445	A	G
chr9	122888831	Zkscan7	T	C
chr9	123151098	Clec3b	G	C
chr9	123151102	Clec3b	T	C
chr9	123977439	Ccr1l1	T	A
chr9	123977442	Ccr1l1	T	C
chr9	124028846	Ccr3	A	G
chr9	124106378	Ccr2	G	A
chr9	124124393	Ccr5	G	T
chr9	124124686	Ccr5	C	G
chr9	124124828	Ccr5	C	T
chr9	124124840	Ccr5	T	C

chr9	124124915	Ccr5	C	T
chr9	124125315	Ccr5	G	T

Table A.10 UM02 Variants Are Most Likely *de novo* Mutations

#CHROM	POS	UM02 Gene_Name	REF	ALT
chr9	4330888	Kbtbd3	G	A
chr9	5321472	Casp4	G	T
chr9	5321474	Casp4	C	T
chr9	6265244	Ddi1	C	A
chr9	6265245	Ddi1	A	C
chr9	7005535	Dync2h1	C	G
chr9	7005537	Dync2h1	A	T
chr9	7142292	Dync2h1	C	A
chr9	7451731	Mmp3	T	A
chr9	7854589	Birc3	C	T
chr9	8099848	AK129341	T	A
chr9	8652980	Trpc6	G	A
chr9	8652983	Trpc6	C	A
chr9	9673889	Cntn5	C	T
chr9	14799616	Mre11a	G	T
chr9	15317709	Cep295	T	A
chr9	15322876	Cep295	G	A
chr9	15332246	Cep295	G	T
chr9	15332247	Cep295	G	A
chr9	15332249	Cep295	T	A
chr9	15332250	Cep295	G	A
chr9	15335628	Cep295	T	A
chr9	15335629	Cep295	C	G
chr9	15335631	Cep295	T	G
chr9	15915300	Fat3	G	T
chr9	15915301	Fat3	G	A
chr9	16004061	Fat3	G	A
chr9	18301055	Chordc1	G	A

chr9	18378626	Naalad2	C	A
chr9	18444701	Mbd3l2	G	T
chr9	18485053	Mbd3l1	C	T
chr9	18815806	Olfr828	C	A
chr9	18945377	Olfr832	C	T
chr9	19566822	Olfr854	C	G
chr9	19808407	Olfr859	C	A
chr9	19808410	Olfr859	T	A
chr9	19808414	Olfr859	T	A
chr9	19920284	Olfr77	C	A
chr9	19920672	Olfr77	G	T
chr9	20100758	Olfr868	G	A
chr9	20260637	Olfr872	T	A
chr9	20772317	Col5a3	G	A
chr9	21008598	Mrpl4	C	G
chr9	21008599	Mrpl4	C	T
chr9	21026384	Icam1	G	T
chr9	24475396	Dpy19l1	G	T
chr9	31329671	Prdm10	G	T
chr9	31410360	Nfrkb	T	A
chr9	31410362	Nfrkb	C	T
chr9	32248578	Arhgap32	A	T
chr9	32256817	Arhgap32	G	A
chr9	32396331	Kcnj1	T	A
chr9	35212357	Srpr	C	A
chr9	37236966	Slc37a2	C	G
chr9	38026348	Olfr883	C	A
chr9	38209851	Olfr893	C	T
chr9	38209853	Olfr893	G	A
chr9	38330177	Olfr25	C	T
chr9	38448913	Olfr902	C	A
chr9	38448916	Olfr902	G	T
chr9	38449381	Olfr902	G	A
chr9	38449613	Olfr902	T	A
chr9	38516769	Olfr908	T	A
chr9	38594267	Olfr913	G	C

chr9	38594268	Olfr913	G	A
chr9	38594539	Olfr913	T	A
chr9	38594542	Olfr913	T	A
chr9	38673201	Olfr918	A	T
chr9	38698174	Olfr919	C	A
chr9	38698347	Olfr919	G	T
chr9	38877485	Olfr926	T	A
chr9	39060253	Olfr937	A	T
chr9	39144322	Olfr27	C	A
chr9	39218142	Olfr944	C	A
chr9	39394624	Olfr951	T	A
chr9	39394625	Olfr951	C	A
chr9	39394627	Olfr951	G	A
chr9	39550590	Olfr958	A	T
chr9	39736878	Olfr150	C	T
chr9	39737007	Olfr150	G	T
chr9	39737008	Olfr150	C	G
chr9	39737010	Olfr150	G	A
chr9	39750434	Olfr967	G	A
chr9	39750436	Olfr967	T	A
chr9	39750446	Olfr967	C	A
chr9	39873683	Olfr972	T	A
chr9	39873684	Olfr972	G	T
chr9	40804979	Hspa8	A	C
chr9	44046656	Thy1	C	T
chr9	44624608	Ddx6	G	T
chr9	44821150	Kmt2a	C	A
chr9	45871952	Rnf214	C	T
chr9	46292175	Bud13	G	A
chr9	48320608	Nxpe2	C	A
chr9	48320661	Nxpe2	T	C
chr9	48326346	Nxpe2	C	G
chr9	48326347	Nxpe2	A	G
chr9	49278697	Gm4894	C	A
chr9	49278699	Gm4894	T	A
chr9	49402245	Drd2	C	T

chr9	49402248	Drd2	C	T
chr9	50800125	Alg9	A	T
chr9	50800127	Alg9	C	A
chr9	51838923	Arhgap20	G	A
chr9	53500796	Atm	C	A
chr9	53527332	Atm	G	A
chr9	53562932	Npat	T	A
chr9	54416736	Dmxl2	T	A
chr9	54428039	Dmxl2	C	A
chr9	55209076	Fbxo22	G	A
chr9	55209101	Fbxo22	G	T
chr9	55209382	Fbxo22	G	A
chr9	55221070	Fbxo22	A	C
chr9	55221151	Fbxo22	G	T
chr9	56257601	Peak1	C	A
chr9	56259768	Peak1	C	A
chr9	56259770	Peak1	C	A
chr9	56259772	Peak1	T	G
chr9	56259773	Peak1	T	A
chr9	56892508	Cspg4	A	T
chr9	57254165	1700017B05Rik	C	T
chr9	58640698	Nptn	A	C
chr9	59780190	Myo9a	C	T
chr9	59780199	Myo9a	A	T
chr9	59832253	Myo9a	C	A
chr9	63524959	Iqch	T	A
chr9	65134598	Igdcc4	G	T
chr9	65475842	Spg21	T	A
chr9	66254595	Dapk2	A	G
chr9	66434141	Herc1	C	A
chr9	66470362	Herc1	G	T
chr9	66955906	Lactb	T	A
chr9	69415810	Ice2	C	T
chr9	69415813	Ice2	T	A
chr9	69997096	Bnip2	A	T
chr9	70429603	Rnf111	T	A

chr9	70429604	Rnf111	C	G
chr9	70443594	Rnf111	G	A
chr9	70586713	Sltm	G	T
chr9	72330989	Zfp280d	C	G
chr9	72330991	Zfp280d	C	A
chr9	72616758	Rfx7	A	T
chr9	72617631	Rfx7	C	T
chr9	72739533	Nedd4	G	T
chr9	72739534	Nedd4	T	A
chr9	72945469	Pygo1	G	C
chr9	73038624	Pigb	T	A
chr9	73932324	Unc13c	T	A
chr9	74218587	Wdr72	C	T
chr9	74889516	Onecut1	G	T
chr9	74889518	Onecut1	A	T
chr9	75388680	Mapk6	T	A
chr9	77792069	Gclc	G	A
chr9	77792071	Gclc	C	A
chr9	77975803	Elovl5	G	A
chr9	78460840	Mto1	T	G
chr9	78460842	Mto1	G	A
chr9	78660940	Cd109	G	C
chr9	78680909	Cd109	G	A
chr9	79693365	Col12a1	T	A
chr9	79703831	Col12a1	C	G
chr9	79703834	Col12a1	C	T
chr9	80281591	Myo6	G	A
chr9	80345592	Impg1	T	A
chr9	85710468	Ibtk	G	A
chr9	85728496	Ibtk	A	T
chr9	85728497	Ibtk	C	A
chr9	85728498	Ibtk	C	A
chr9	85743552	Ibtk	C	A
chr9	85743692	Ibtk	C	T
chr9	86512594	Dopey1	G	C
chr9	87193680	Cep162	C	T

chr9	87248444	Cep162	C	G
chr9	87248482	Cep162	A	T
chr9	87248485	Cep162	G	T
chr9	88398364	Snx14	G	T
chr9	88398366	Snx14	G	C
chr9	88476906	Syncrip	C	A
chr9	88476908	Syncrip	C	G
chr9	89596561	AF529169	G	T
chr9	89602539	AF529169	C	A
chr9	89944805	Rasgrf1	A	C
chr9	89944806	Rasgrf1	C	A
chr9	92291100	Plscr2	T	G
chr9	92595332	Plod2	T	C
chr9	95495933	U2surp	G	T
chr9	95977862	Xrn1	A	T
chr9	95981924	Xrn1	G	T
chr9	96038724	Xrn1	C	T
chr9	96038725	Xrn1	A	G
chr9	96340306	Atp1b3	C	T
chr9	96688188	Zbtb38	G	A
chr9	98581193	Copb2	T	A
chr9	98581195	Copb2	C	A
chr9	98903746	7420426K07Rik	C	A
chr9	99073571	Pik3cb	T	C
chr9	99579446	Dbr1	T	C
chr9	99579449	Dbr1	G	T
chr9	99717062	Cldn18	A	T
chr9	100880875	Stag1	C	A
chr9	101211583	Ppp2r3a	C	A
chr9	101929294	Ephb1	A	T
chr9	103328412	Topbp1	G	C
chr9	104005513	Nphp3	T	A
chr9	105939724	Col6a5	C	A
chr9	106295203	Poc1a	T	A
chr9	106485138	Rrp9	G	A
chr9	106858277	Vprbp	C	A

chr9	106858279	Vprbp	G	A
chr9	106978744	Dock3	G	T
chr9	107765487	Rbm5	T	A
chr9	108405410	Ccdc36	T	A
chr9	108563279	Impdh2	G	A
chr9	108579833	P4htm	T	C
chr9	108796297	Ip6k2	T	A
chr9	108796298	Ip6k2	T	A
chr9	108970508	Col7a1	G	T
chr9	109089489	Ccdc51	G	A
chr9	109891756	Cdc25a	T	A
chr9	110517624	Kif9	C	A
chr9	110549405	Setd2	C	A
chr9	110592569	Setd2	C	T
chr9	111390746	Trank1	G	T
chr9	111572380	Stac	C	T
chr9	115246047	Stt3b	T	A
chr9	122861850	Zfp445	C	G
chr9	122861902	Zfp445	A	G
chr9	122888831	Zkscan7	T	C
chr9	122948015	1110059G10Rik	G	A
chr9	122948018	1110059G10Rik	C	A
chr9	123151098	Clec3b	G	C
chr9	123151102	Clec3b	T	C
chr9	123569996	Sacm1l	G	T
chr9	124028800	Ccr3	G	A
chr9	124028846	Ccr3	A	G
chr9	124029090	Ccr3	C	A
chr9	124106378	Ccr2	G	A
chr9	124124393	Ccr5	G	T
chr9	124124686	Ccr5	C	G
chr9	124124828	Ccr5	C	T
chr9	124124840	Ccr5	T	C
chr9	124124915	Ccr5	C	T
chr9	124125315	Ccr5	G	T

Table A.11 UM03 Variants Are Most Likely *de novo* Mutations

		UM03		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	55209076	Fbxo22	G	A
chr9	55209101	Fbxo22	G	T
chr9	55209382	Fbxo22	G	A
chr9	55221070	Fbxo22	A	C
chr9	55221151	Fbxo22	G	T
chr9	75388680	Mapk6	T	A
chr9	122861850	Zfp445	C	G
chr9	122861902	Zfp445	A	G
chr9	122888831	Zkscan7	T	C
chr9	123151098	Clec3b	G	C
chr9	123151102	Clec3b	T	C
chr9	124028800	Ccr3	G	A
chr9	124028846	Ccr3	A	G
chr9	124106378	Ccr2	G	A
chr9	124124393	Ccr5	G	T
chr9	124124686	Ccr5	C	G
chr9	124124828	Ccr5	C	T
chr9	124124840	Ccr5	T	C
chr9	124124915	Ccr5	C	T
chr9	124125315	Ccr5	G	T

Table A.12 UM11 Variants Are Most Likely *de novo* Mutations

		UM11		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	4384050	Msantd4	G	A
chr9	13620666	Maml2	C	G
chr9	16375123	Fat3	T	G
chr9	19783342	Olfr58	G	T
chr9	19783344	Olfr58	C	G
chr9	19783346	Olfr58	T	A

chr9	19783347	Olfr58	G	A
chr9	22073489	Ecsit	C	G
chr9	37804144	Olfr876	C	A
chr9	37897523	Olfr145	G	A
chr9	37993020	Olfr881	A	G
chr9	38025884	Olfr883	C	A
chr9	38047249	Olfr884	T	A
chr9	38330231	Olfr25	C	G
chr9	38368163	Olfr899	G	A
chr9	38368165	Olfr899	G	C
chr9	38516769	Olfr908	T	G
chr9	38698217	Olfr919	G	T
chr9	38698223	Olfr919	G	T
chr9	38698224	Olfr919	G	C
chr9	38698225	Olfr919	A	C
chr9	38698231	Olfr919	A	C
chr9	38698235	Olfr919	C	T
chr9	38734706	Vwa5a	T	A
chr9	38877245	Olfr926	C	T
chr9	39217779	Olfr944	T	A
chr9	39669141	Olfr963	T	C
chr9	39750446	Olfr967	C	A
chr9	39750547	Olfr967	C	A
chr9	39750587	Olfr967	C	A
chr9	39839757	Olfr971	G	T
chr9	45176520	Tmprss4	C	A
chr9	49507015	Ncam1	T	A
chr9	50754536	Cryab	T	A
chr9	50754540	Cryab	T	A
chr9	50870073	Ppp2r1b	T	A
chr9	55209076	Fbxo22	G	A
chr9	55209101	Fbxo22	G	T
chr9	55221070	Fbxo22	A	C
chr9	58025169	Cyp11a1	G	A
chr9	66133477	Fam96a	A	T
chr9	66133479	Fam96a	A	C

chr9	75347972	Bcl2l10	G	A
chr9	78157903	Ick	A	T
chr9	78478834	Eef1a1	C	T
chr9	78480516	Eef1a1	C	A
chr9	79700278	Col12a1	G	T
chr9	104176665	Dnajc13	A	T
chr9	109192770	Fbxw13	G	T
chr9	109494760	Fbxw19	C	G
chr9	109494761	Fbxw19	T	C
chr9	109722093	Fbxw26	G	C
chr9	113915213	Clasp2	G	T
chr9	114430321	Glb1	T	G
chr9	114430325	Glb1	T	G
chr9	116109988	Tgfbr2	T	A
chr9	118115340	Cmc1	C	A
chr9	119609842	Scn10a	A	C
chr9	122856819	Zfp445	A	T
chr9	123183722	Cdcp1	C	G
chr9	123183723	Cdcp1	T	G
chr9	124125017	Ccr5	T	G

Table A.13 UM12 Variants Are Most Likely *de novo* Mutations

		UM12		
#CHROM	**POS**	**Gene_Name**	**REF**	**ALT**
chr9	3458760	Cwf19l2	G	A
chr9	4472124	Gria4	G	A
chr9	4472125	Gria4	C	A
chr9	6265403	Ddi1	C	A
chr9	7447622	Mmp3	T	A
chr9	7447625	Mmp3	A	C
chr9	15084266	Hephl1	A	G
chr9	18875341	Olfr830	G	C
chr9	18875913	Olfr830	C	A
chr9	18875920	Olfr830	A	T

chr9	19035320	Olfr835	T	A
chr9	20101001	Olfr868	G	A
chr9	20285896	Olfr39	C	T
chr9	37855457	Olfr877	C	A
chr9	37897665	Olfr145	A	T
chr9	37897668	Olfr145	A	G
chr9	37897670	Olfr145	A	T
chr9	37919430	Olfr878	T	A
chr9	38084842	Olfr887	G	A
chr9	38085333	Olfr887	A	T
chr9	38085334	Olfr887	C	T
chr9	38109181	Olfr888	G	C
chr9	38109184	Olfr888	G	A
chr9	38109185	Olfr888	C	A
chr9	38109188	Olfr888	G	T
chr9	38109189	Olfr888	C	G
chr9	38209811	Olfr893	T	A
chr9	38254085	Olfr143	A	G
chr9	38254124	Olfr143	G	A
chr9	38329601	Olfr25	G	A
chr9	38329602	Olfr25	C	T
chr9	38349094	Olfr898	A	C
chr9	38403191	Olfr147	C	A
chr9	38403197	Olfr147	C	A
chr9	38464190	Olfr904	G	A
chr9	38594542	Olfr913	T	A
chr9	38815802	Olfr922	C	A
chr9	38848338	Olfr924	A	G
chr9	38848339	Olfr924	C	A
chr9	38994847	Olfr935	T	G
chr9	39060370	Olfr937	C	G
chr9	39060371	Olfr937	C	A
chr9	39060397	Olfr937	T	C
chr9	39184401	Olfr943	C	A
chr9	39217779	Olfr944	T	A
chr9	39257835	Olfr945	G	A

chr9	39462253	Olfr954	G	A
chr9	39511397	Olfr957	C	T
chr9	39511454	Olfr957	T	G
chr9	39511459	Olfr957	A	T
chr9	39623431	Olfr960	T	G
chr9	39646810	Olfr961	T	A
chr9	39646812	Olfr961	T	A
chr9	39646813	Olfr961	T	G
chr9	39646814	Olfr961	T	A
chr9	39646815	Olfr961	T	C
chr9	39795576	Olfr969	C	A
chr9	39819851	Olfr970	C	A
chr9	39820371	Olfr970	A	C
chr9	39820455	Olfr970	A	T
chr9	40802745	Hspa8	A	T
chr9	40804680	Hspa8	G	T
chr9	45911080	Pcsk7	T	G
chr9	54416736	Dmxl2	T	A
chr9	54884024	Ireb2	T	C
chr9	54906654	Ireb2	C	A
chr9	55209076	Fbxo22	G	A
chr9	55209360	Fbxo22	C	T
chr9	55209382	Fbxo22	G	A
chr9	55215004	Fbxo22	A	T
chr9	55221070	Fbxo22	A	C
chr9	55221151	Fbxo22	G	T
chr9	56970309	Snupn	T	C
chr9	58025169	Cyp11a1	G	A
chr9	58878153	Neo1	G	T
chr9	66501258	Herc1	C	G
chr9	70410215	Ccnb2	G	A
chr9	72615157	Rfx7	T	A
chr9	72615159	Rfx7	C	A
chr9	73699120	Unc13c	C	A
chr9	75388680	Mapk6	T	A
chr9	75393362	Mapk6	T	G

chr9	78328548	Omt2b	T	C
chr9	79680367	Col12a1	C	A
chr9	79680368	Col12a1	C	T
chr9	80262353	Myo6	G	T
chr9	85710468	Ibtk	G	A
chr9	85710469	Ibtk	C	A
chr9	85710470	Ibtk	C	A
chr9	85710471	Ibtk	T	A
chr9	85743479	Ibtk	C	T
chr9	87227188	Cep162	G	T
chr9	95464363	U2surp	C	T
chr9	103320639	Topbp1	G	T
chr9	103338288	Topbp1	G	T
chr9	104018241	Nphp3	A	G
chr9	104024628	Nphp3	A	C
chr9	104024641	Nphp3	C	T
chr9	104024658	Nphp3	A	G
chr9	104029997	Nphp3	A	G
chr9	104031987	Nphp3	G	A
chr9	104075894	Acad11	A	C
chr9	104203437	Dnajc13	C	T
chr9	106026522	Col6a4	C	T
chr9	110549496	Setd2	A	C
chr9	113958852	Ubp1	A	T
chr9	118689396	Itga9	C	A
chr9	122857512	Zfp445	A	G
chr9	122861850	Zfp445	C	G
chr9	122861902	Zfp445	A	G
chr9	124124699	Ccr5	G	A

Table A.14 Possible *de novo* Mutations Identified in Step 7 Are Common to All LL Samples and Are Also Present in the Typical Samples

#CHROM	POS	Gene_Name	REF	ALT
chr9	55209076	Fbxo22	G	A
chr9	55221070	Fbxo22	A	C

REFERENCES

1. De Wit P, Pespeni MH, Ladner JT, Barshis DJ, Seneca F, Jaris H, Overgaard Therkildsen N, Morikawa M and Palumbi SR. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources, 12*, 1058-1067.

2. Noah's Ark Children's Services Resource Unit. (2009). Infantile Spinal Muscular Atrophy

3. Sheets, Kayla Mandel, M.S. (2014). Spinal muscular atrophy. *Salem Press Encyclopedia of Health*

4. Thomas W Prior, Barry S Russman. (2013). GeneReviews: Spinal Muscular Atrophy

5. Sophie Schbath, Véronique Martin, Matthias Zytnicki, Julien Fayolle, Valentin Loux, and Jean-François Gibrat. (2012). Mapping Reads on a Genomic Sequence: An Algorithmic Overview and a Practical Comparative Analysis. *J Comput Biol, 19(6),* 796–813.

6. Spinal muscular atrophy. (2013). *The Genetics Home Reference*

7. AHM Lai, ES Tan, HY Law, CS Yoon, ISL Ng. (2005). SMN1 Deletions Among Singaporean Patients With Spinal Muscular Atrophy. *Annals Academy of Medicine. 34*

8. Zerres K, Rudnik-Schoneborn S, Forkert R, Wirth B. (1995). Genetic basis of adult-onset spinal muscular atrophy. *Lancet. 346,* 1162.

9. Yang Z, Yoder AD. (1999). Estimation of the transition/transversion rate bias and species sampling. *J Mol Evol. 48(3),* 274-83.

10. M Taketo, A C Schroeder, L E Mobraaten, K B Gunning, G Hanten, R R Fox, T H Roderick, C L Stewart, F Lilly, C T Hansen. (1991). FVB/N: an inbred mouse strain preferable for transgenic analyses. *Proc Natl Acad Sci U S A. 88(6),* 2065-2069.

11. Hahnen E, Schonling J, Rudnik-Schoneborn S, Raschke H, Zerres K, Wirth B. (1997). Missense mutations in exon 6 of the survival motor neuron gene in patients with spinal muscular atrophy (SMA). *Hum Mol Genet. 6,* 821–5.