Duplication Tag SNP g.27134T>G should not be considered diagnostic of SMA carrier status

Dan Davison, PhD; K. Eerik Kaseniit, MEng; Imran S. Haque, PhD

Counsyl

South San Francisco, California

Background

Spinal muscular atrophy (SMA) is a severe neuromuscular disease that is the second most common fatal autosomal recessive disorder[1]. The most common SMA carrier alleles involve a deletion in SMN1[1]. However, duplication alleles also exist, and common methods of testing (e.g., quantitative PCR) measure total dosage and are therefore unable to distinguish a 2+0 carrier from a 1+1 non-carrier. Luo et al.[2] (2014) describe a tag SNP g.27134T>G found to be in linkage with the duplication allele. Following the recommendations of this publication, clinical diagnostic laboratories have begun to use the presence of this tag SNP in 2-copy individuals as diagnostic of SMA carrier status in Ashkenazi Jewish (AJ) and Asian populations.

Methods

We developed a full-likelihood Bayesian model for the CN and SNP data and for post-test risk. The model for the CN and SNP count data has the structure of a multinomial model, in which the cell probabilities depend on the (unknown) SNP + copy number haplotype frequencies in a given population.

We calculated the residual risk that a CN=2 individual is an SMA carrier conditional on their SNP genotype by sampling from the posterior distribution of the haplotype frequencies, and using these samples to construct a sample from the posterior distribution of residual risk values. This procedure propagates all uncertainty regarding population allele frequencies and SNP/copy-number correlations through to the final residual risk calculation. We use rejection sampling to sample from posterior probability distributions that are not analytically tractable.

We sequenced 12,089 individuals, randomly selected without regard to ethnicity or SMA genotype/phenotype, to determine *SMN1* CN and g.27134T>G SNP genotype. We computed posterior residual risk values for the original Luo et al.[2] data set, and for the new Counsyl data set.

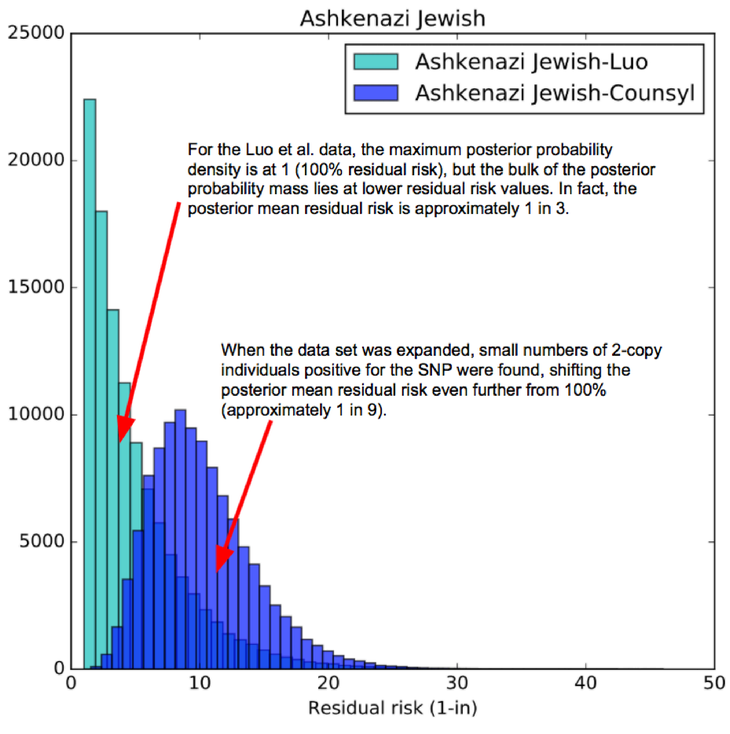
Population	CN	Luo et al. data			Counsyl data		
		AA	Aa	aa	AA	Aa	aa
Ashkenazi Jewish	1	0	0	0	10	0	0
	2	315	**0**	0	640	**4**	0
	3	21	86	4	55	183	52
	4	1	23	0	4	19	43
(Eastern) Asian	1	2	0	0	11	0	0
	2	222	**0**	0	628	**1**	0
	3	20	2	0	45	1	0
	4	1	1	0	0	0	0
Caucasian	1	12	0	0	181	0	0
	2	413	2	0	6871	43	2
	3	23	4	0	456	151	21
	4	2	2	0	20	14	15
Hispanic	1	1	0	0	18	0	0
	2	206	12	1	783	16	1
	3	20	20	0	61	37	6
	4	0	2	0	2	1	0

Raw counts from the original study (Luo et al. 2014) and from Counsyl data.

Within each population, the counts are divided into distinct *SMN1* copy number categories (rows labeled 1, 2, 3, 4). Within each such copy number category, the counts are divided into 3 columns according to the tag SNP genotype (AA=negative, Aa=heterozygous positive, aa=homozygous positive)

Results

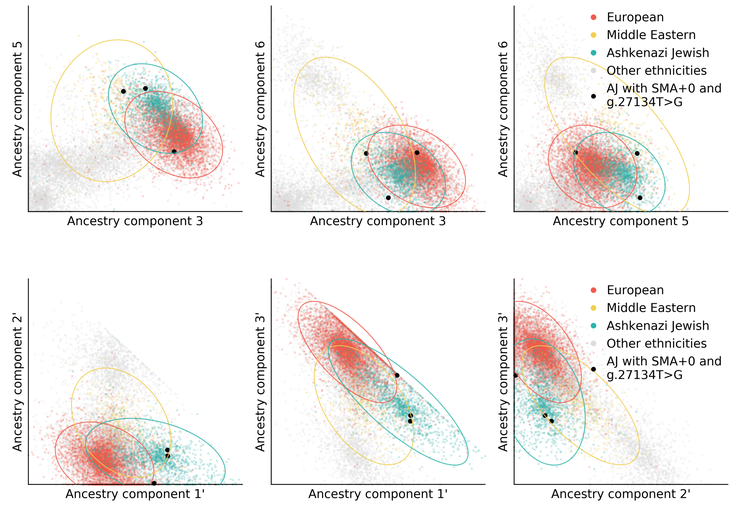
- Estimating the posterior distribution over model parameters shows that maximum likelihood / least-squares estimation (ML/LS) can lead to erroneous calculation of residual risk, with ML/LS estimates being either substantially too high or too low. For example, although ML/LS residual risk for AJs in the Luo et al. data is 100% (1 in 1), the posterior mean value of this risk is much lower: 1 in 2.9. Similarly, for Asian individuals the ML/LS estimate is 100% (1 in 1), whereas the posterior mean is 1 in 7.9.
- In contrast to the Luo et al. data, we found a small number of 2-copy AJ individuals positive for the tag SNP, whose presence decreases the inferred linkage between the duplication allele and the SNP. These individuals had genetic ancestry typical of self-reported AJ individuals.
- A diverse test population demonstrated heterogeneity of tag SNP performance among Asian populations: while self-reported Southeast Asians had statistics similar to the Luo et al. "Asian" population, among both East and South Asians the inferred linkage between the tag SNP and the *SMN1* duplication is substantially lower, conveying much smaller residual risk.



Although the posterior mode is at 100% for the Luo et al. data, it is not appropriate to treat the tag SNP as diagnostic for SMA carrier status in the Ashkenazi Jewish population.

The histograms illustrate the posterior distribution of Ashkenazi Jewish residual risk of being an SMA carrier after testing positive for g.27134T>G (Luo et al. data | Counsyl data). Each histogram comprises 100,000 samples simulated from the posterior distribution of the residual risk. Notice that for the Luo et al. data, while the probability distribution peaks at a residual risk of 100% (1 in 1), there is substantial probability weight away from this mode. In fact, the posterior mean is 1 in 2.9. The conclusion that the residual risk is substantially lower than 100% is even stronger for the Counsyl data (posterior mean residual risk 1 in 8.7).

		Risk (1-in-X) of being an SMA carrier given SNP positive 2-copy genotype		
Population	Data Source	MLE	Posterior Mean	95% CI
African American	Luo et al.	35.2	35.2	25.6 - 48.2
	Counsyl	38.7	**38.7**	33.2 - 45.2
Ashkenazi Jewish	Luo et al.	1	2.9	1.1 - 15.8
	Counsyl	8.7	8.7	4.1 - 19.7
Caucasian	Luo et al.	27	**23.1**	7.1 - 102.1
Northern Europe	Counsyl	31.2	31.1	24.1 - 40.3
Southern Europe	Counsyl	103.5	40.3	8.9 - 404.1
Asian	Luo et al.	1	7.9	1.5 - 101.9
Eastern Asia	Counsyl	122.9	45.5	10.1 - 454.7
Southeast Asia	Counsyl	1	10	1.7 - 149.9
South Asia	Counsyl	2229.5	51.7	10.8 - 665.2
Hispanic	Luo et al.	147.8	137.8	77.0 - 259.9
	Counsyl	82.4	81	51.5 - 132.1
Middle East	Counsyl	19.4	16.7	4.0 - 109.9



The 2-copy individuals who self-reported as Ashkenazi Jewish and are positive for the tag SNP are within the normal range of Ashkenazi Jewish genetic ancestry.

The top panel shows global genetic ancestry (amount of ancestry shared with 7 hypothetical ancestral populations defined using reference populations across the world), highlighting 3 individuals who are both 2-copy and positive for the SNP tag (small black circles). Individuals self-reporting as European, Middle Eastern, or Ashkenazi Jewish are also shown.

The bottom panel shows the same cohort, analyzed for local genetic ancestry shared with 3 hypothetical ancestral populations defined using only European, Middle Eastern, and Ashkenazi Jewish reference populations, highlighting the same 3 individuals.

Conclusions

- Uncertainty in population haplotype frequency estimates should be propagated forwards into residual risk calculations using a standard Bayesian probability framework. The original study's conclusions were based on least-squares/maximum likelihood, but using these analysis methods leads to the conclusion in certain populations that a proband testing positive for the tag SNP is certainly a carrier, when in fact it is still more probable that they are not.
- Phasing experiments could resolve carrier status in individual cases and would improve understanding of the implications of the tag SNP for carrier status.

REFERENCES: 1. Sugarman et al. (2012) Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy. Eur J Hum Genet 20(1):27- 32. | **2.** Luo et al. (2014) An Ashkenazi Jewish SMN1 haplotype specific to duplication alleles improves pan-ethnic carrier screening for spinal muscular atrophy. Genet Med 16(2):149-156

View all posters and research at
research.counsyl.com