SMALL SEGMENT ANALYSIS - JULIAN LAND'S PROCEDURE

After Andrew Millard partially corrected my lack of knowledge of reproductive biology in May 2025, my only way forward was to complete the analysis of all 1963 3cM segments generated from my 12-relative set by GEDmatch (at P = 3). From this my procedure for using small segments to prove family branch connections can now be documented. The following steps will help others do the same thing.

STEP 1

Learn the key - we are looking for rare segment boundary coincidences (RSBCs). These occur when 2 independent pairs of relatives have a coincident boundary (left or right). I found 46 of them - 25 on the left boundary, 21 on the right.

Because the number of these occurrences greatly exceeds the number expected at random (see **Appendix 1**), each RSBC is significant and thus conveys some genetic information. If pair A&B forms a RSBC with pair C&D, then either A matches C or D, or B matches C or D. Of course, A&B and C&D might also be genuine matches. Better genetic information can be obtained as follows.

STEP 2

Put all matches in order, using the left boundary, with each chromosome on its own (Excel) worksheet. Find each RSBC and colour it the same colour whichever chromosome it sits in. Generally review each RSBC – noting where (as in the case of location 752721 on Cr1) some hard genetic boundary is evident leaving to a large number of matches terminating there in a non-random fashion. Another such location (135656 on Cr10) was similarly dismissed.

STEP 3

Inspect the matches around each RSBC, looking for evidence that the RSBC will yield more genetic information. In my case I found that 14 RSBCs had 4 possible interpretations, as explained above. This was after – sometimes after an iteration through the entire set of connections proven to that point – I found that one of the 4 possible interpretations had already been proven in which case no conclusion was possible, even if one of the 2 matches had been proven. In this way 16 RSBCs were dismissed. So 14 of the 1 in 4 situations remained but these were not used.

STEP 4

Some of the remaining RSBCs show just 3 possibilities eg A matches C or D. This can be useful where C and D have a known genetic or genealogical link though here we dismissed one of these cases but the remaining 2 was able to be used.

But much more progress can be made with the RSBCs which are amenable to logical deduction.

STEP 5

Here strict logic is required to reject any match around a RBSC which could be a false positive, whilst using any evidence to the contrary, usually in the form of linked 3-person Segment Boundary Coincidences (3pSBCs).

In my case, I was able to prove 12 genetic matches, quite enough to confirm connection between my family branches.

STEP 6

Display the results as in **Diagram 1,** so that the level of family connection between the 12 (in my case) relatives can be seen. The Diagram shows 11 are related via 12 proven family links. Missing are 54 links for there are 66 possible pairings. **Diagram 1** could also show the Step 3 and Step 4 cases.

Most pairs can be linked via 2 connections. Some relatives play a key role with more connections than others – Ju, JG and Je have 4 connections and , with SW and TP having 3. Those with the fewest connections are RF and RM having 2 connections, E and B having 1 connection each and A, MD and J none.

Diagram 2 shows the 10 connections resolved through commercial DNA matching, only 1 of which duplicates our results. **Diagram 3** shows the 14 known matches in a branch not picked up by our method though 4 are pocked up by commercial DNA matching.

CONCLUDING COMMENT

It is certainly true that small segments are a challenge for those who have to work with them. We describe a method of dealing with this challenge, starting with the set of Rare Segment Boundary Coincidences (RSBCs) which occurs amid any large set of small 3cM matches generated from a set of relatives.

Previous work indicated that a sample of relatives generated 5% more 3cM matches than the same-sized random sample. Here our 46 RSBCs comprise 2-3% of our 1963 matches. The number of proven family matches was found to be about 1% of 1963.

Here, our procedure finds 12 matches, 1 of which was known through standard commercial DNA testing. Our procedure also misses 14 intra-branch matches known through genealogy, but commercial matching misses 10 of these. Our procedure complements both genealogy and commercial autosomal matching, thus amply confirms inter-branch connection.

Julian Land

23 October 2025

APPENDIX 1 - PROBABILITY

The total number of coincidences expected at random involving 4 people from 12 in a set of N matches over each of 22 chromosomes would be 22*N*(N-1)100,000 = 1.7. We found 25 RSBCs (left) and 21 RSBCs (right). The 3-person Segment Boundary Coincidences (SBCs) occur more frequently. We suggest the excess incidence in both categories reflects the impact of family.

In this estimate of random occurrences, there were about (N=) 89 segment matches found per chromosome at the settings used (Qmatch 3cM P=3), and there are about 50,000 SNPs per chromosome.

The 3-person version involves one person occurring twice but not in the same match. The number of 4 person matches from M is:

$$M*(M - 1)*(M - 2)*(M - 3)/24 = 495 \text{ when } M = 12$$

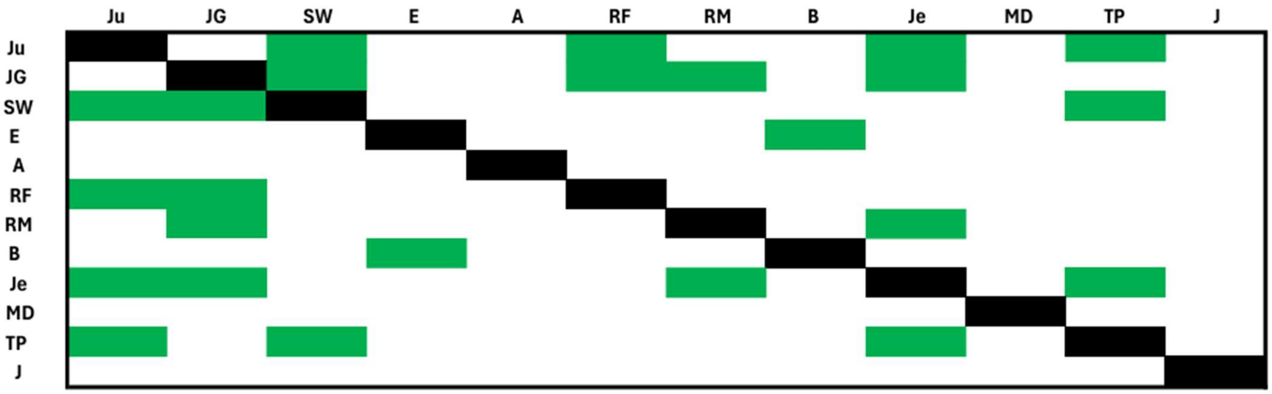
The number of 3-person matches from M is:

$$2/3*3*M*(M - 1)*(M - 2)/6 = 440 \text{ when } M = 12$$

Of course, this does not mean that the number of 3-person SBCs should be smaller than the number of RSBCs. In fact it is much larger, due to the fact that the same person across a pair of matches brings in that person’s genetic history. The large number of SBCs thus contains family data but of course not all of it from the target family. One’s initial temptation to dismiss 3pSBCs should be resisted, because they play a role in distilling the connections lying in the RSBCs.

DIAGRAM 1

Here we show the 12 family matches generated by 46 RSBCs.

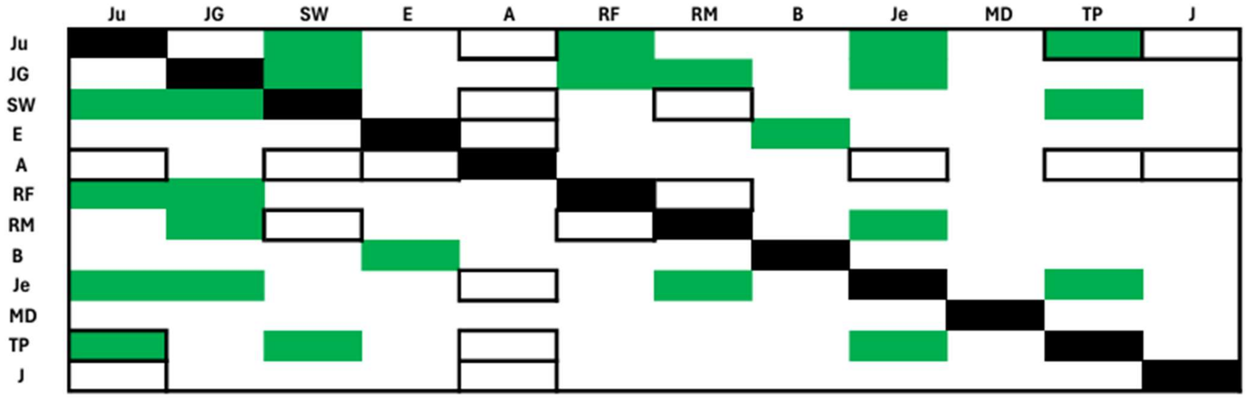


	Ju	JG	SW	E	A	RF	RM	B	Je	MD	TP	J
Ju	⬛		🟩			🟩			🟩		🟩	
JG		⬛	🟩			🟩	🟩		🟩			
SW	🟩	🟩	⬛								🟩	
E				⬛				🟩				
A					⬛							
RF	🟩	🟩				⬛						
RM		🟩					⬛		🟩			
B				🟩				⬛				
Je	🟩	🟩					🟩		⬛		🟩	
MD										⬛		
TP	🟩		🟩						🟩		⬛	
J												⬛

We have not included the 1 in 4 possibilities, nor the other uncertain cases.

DIAGRAM 2

Here, the matches in Diagram 1 are augmented by the 10 cases proven by Qmatch at 7cM with P=3. These 10 cases are outlined. The conditions are strict enough to disallow B’s matches with Je and TP found on Ancestry.com.



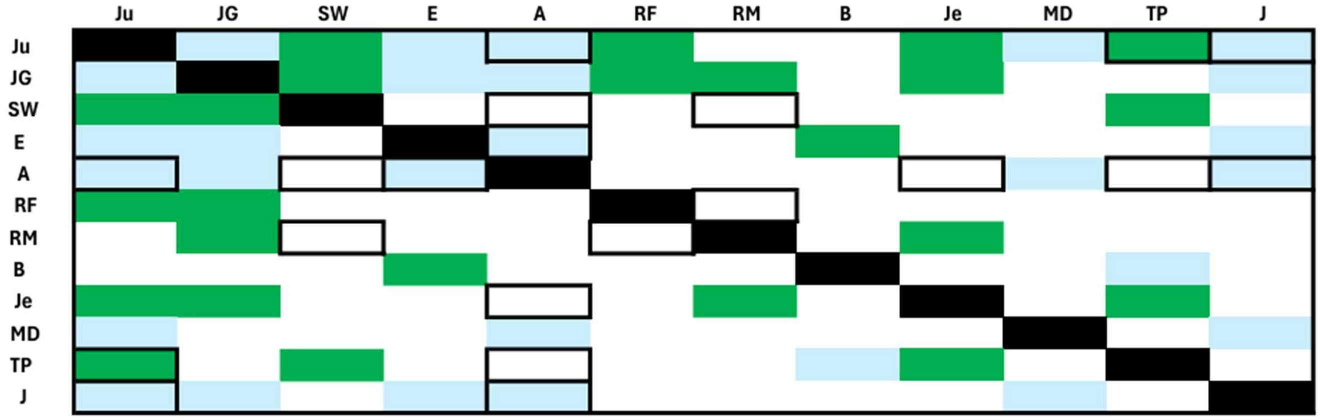
	Ju	JG	SW	E	A	RF	RM	B	Je	MD	TP	J
Ju	■		🟩		☐	🟩			🟩		🟩☐	☐
JG		■	🟩			🟩	🟩		🟩			
SW	🟩	🟩	■		☐		☐				🟩	
E				■	☐			🟩				
A	☐		☐	☐	■				☐		☐	☐
RF	🟩	🟩				■	☐					
RM		🟩	☐			☐	■		🟩			
B				🟩				■				
Je	🟩	🟩			☐		🟩		■		🟩	
MD										■		
TP	🟩☐		🟩		☐				🟩		■	
J	☐				☐							■

There is overlap only at Ju/TP, so there is complementarity. Both techniques show connection across family branches.

Remarkably all the commercial matches arose from the 5 additional family members (A, RF, RM, TP, and J) and it was the absence of commercial matches among our 7 original family members (Ju, JG, SW, E, B, Je and MD) which induced us to try a new approach.

DIAGRAM 3

Here we show 14 intra-branch matches not found by our small match procedure. The commercial matches however pick up only 4 of them.



In our situation we are apparently at the limits of technology, for we are unable to prove 10 intra-branch matches either by commercial matches or our small match procedure.