SMALL SEGMENT ANALYSIS - JULIAN LAND'S PROCEDURE

After Andrew Millard taught me some reproductive biology in May 2025, my only way forward was to carry on to complete the analysis of all 1962 3cM segments generated from my 12-relative set by GEDmatch (at P = 3). From this I completed my procedure for using small segments to prove family branch connections. 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 40 of them - 22 on the left boundary, 18 on the right.

Because the number of these occurrences greatly exceeds the number expected at random (0.12 in each case - 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. 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.

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 25 RSBCs could not be improved. That is, each had 4 possible interpretations, as explained above.

STEP 4

Of the remaining RSBCs, improvements can be made. In my case, 7 of the remaining 15 RSBCs could be improved to show just 2 possibilities eg A matches C or D. This can be useful where C and D have a known genetic or genealogical link.

But much more progress can be made.

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 other (linked) coincidences.

In my case, in addition to the 7 partial improvements mentioned above, I was able to prove 15 genetic matches, quite enough to confirm connection between my family branches.

STEP 6

Use ASBs (defined in **Appendix 2**) to augment our tally of proven family matches if required. **Diagram 1** illustrates this step, not strictly required in this case.

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. Then we can use ASBs (defined in **Appendix 2**) to augment the proven family matches we find using RSBCs.

Previous work indicated that a sample of relatives generated 5% more 3cM matches than the same-sized random sample. Here our 40 RSBCs and 17 ABCs together comprise 3% of our 1962 matches. The number of proven family matches was found to be about 2% of 1962.

Julian Land

4 August 2025

APPENDIX 1 - PROBABILITY

The total number of lefthand RSBCs or righthand RSBCs expected at random would be 66*30*3/50,000 = 0.12 - but we found 21 (left) and 17 (right) a vast excess which presumably reflects the impact of family. In this estimate of random occurrences, there are 66 different relative pairs, about 30 matching segments per pair, only about 3 segments from the other pair can match a given segment from a pair of relatives, and there are about 50,000 SNPs per chromosome (only 1 of which will match a boundary of both segments).

APPENDIX 2 - ASBs

Abutting segment boundaries (ASBs) we found to be as common as each type of RSBC (17 ASB locations in our set of 1962 matches). Their random occurrence frequency is the same as for each type of RSBC. By analogy, the ASBs likewise promise a family connection.

We found 14 ASBs containing 3 relatives where a common relative matches 2 nonmatching relatives ie 2 proven family links each. 4 ASBs contained 4 relatives and show a family match either side of the ASB - again 2 more family links each. Our 17 ASB locations generated 36 family matches - a few of which overlap.

The only thing intuitively obvious about the family relevance of ASBs is that the 3-relative version is found more than 3 times as often as the 4-relative version (greatly exceeding our expectation of less than 1/4 times). This excess suggests we should use only the 3-relative version.

With my 12-relative set, and using 3-relative ASBs to augment the family matches found with RSBCs, I arrived at **Diagram 1**. While we start our small-segment procedure with RSBCs, it may be more practicable to reverse the procedure by starting with ASBs and augmenting with RSBCs.

DIAGRAM 1

Here we show the family matches generated by RSBCs (red) and the additional family matches generated by 3-relative ASBs (yellow). Overlaps between the 2 types are shown in orange.

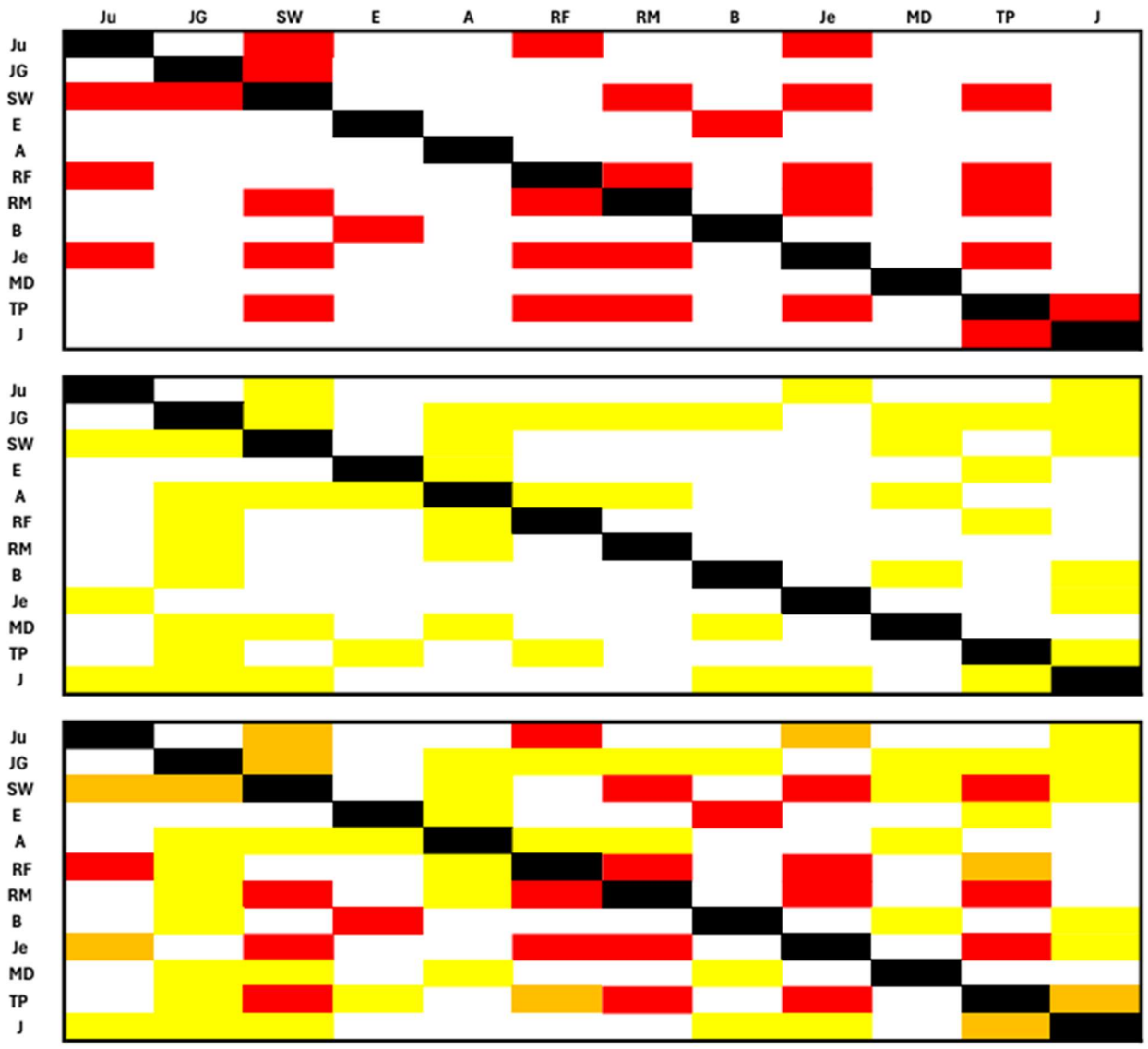
Notes:

1. The 1st chart shows 15 proven matches derived from RSBCs after allowing for 2 overlaps.
2. The 2nd chart shows 24 proven matches derived from 3-rel ASBs after allowing for 4 overlaps.
3. The 3rd chart combines the first 2 charts, showing 34 proven matches with 5 overlaps (shown).
4. The RSBC proven matches rely on many matches by SW, Je and TP.
5. The ABS proven matches rely on many matches by JG, A, and J.
6. Those with fewer matches - Ju, MD, E and B - have well known genealogies within a family branch.
7. The match patterns are complementary with only 5 overlaps between RSBC and ASB proven matches.

from 3-relative ASBs

both 3-relative ASBs and RSBCs

from RSBCs


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