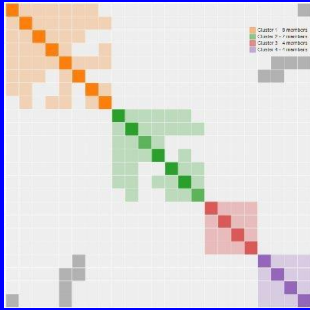# DNA Special Interest Group

Genetic Affairs

- Clusters & Segments

BIFHSGO – DNA SIG

June 5, 2021

R.E. (Bob) Butler

rebutler@storm.ca

Clusters & Segments
Presentation Overview

- ☐ Total Match Clusters   Demo & 8 Slides

- ☐ Segment Clusters   7

- ☐ DNA Pileups   6

- ☐ Triangulation & Final Clusters   4

- ☐ Genetic Affairs Summary   2
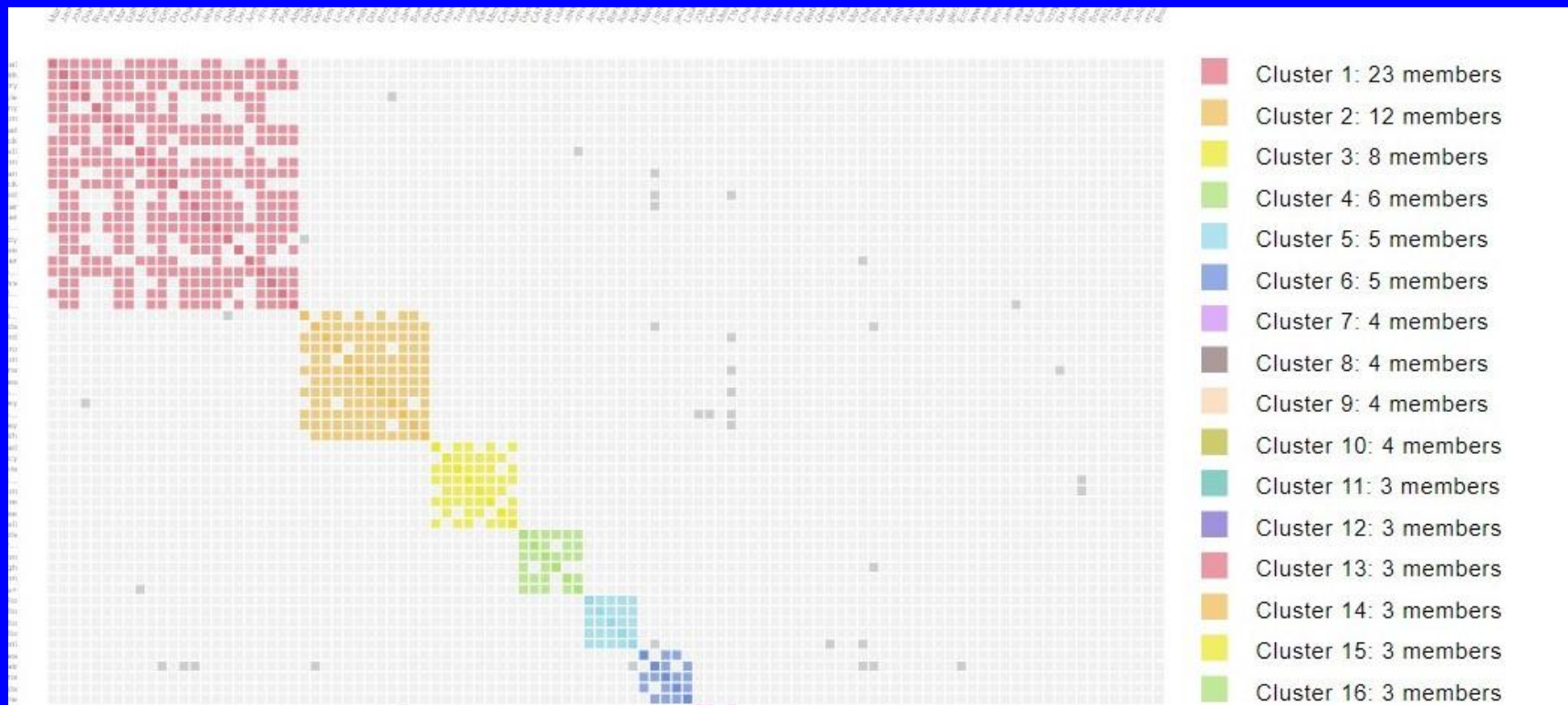
Match Cluster

Cluster = Part of an Organized Match List

- Group People who match each other
- Probably a branch of your tree
- Clusters help focus further research

My Heritage Name	Cheryl	Harry	Trish	Virginia	Karen	Michael	Carolyn	Marie	REB Total Shared cM	People in Tree
Cheryl	-		39	33		23		15	40.72	8
Harry		-	101	2438	29	47	38		53.46	86
Trish	39	101	-	82	15	33		53	70.92	
Virginia	33	2438	82	-	32	21	18	17	48.67	152
Karen		29	15	32	-	34	37		30.18	1
Michael	23	47	33	21	34	-		39	35.66	1
Carolyn		38		18	37		-	26	44.18	17
Marie	15		53	17		39	26	-	51.17	3

Cluster 3

Genetic Affairs - MyHeritage
AutoCluster Sort Display & Cluster List 13

	Cluster 1: 23 members
	Cluster 2: 12 members
	Cluster 3: 8 members
	Cluster 4: 6 members
	Cluster 5: 5 members
	Cluster 6: 5 members
	Cluster 7: 4 members
	Cluster 8: 4 members
	Cluster 9: 4 members
	Cluster 10: 4 members
	Cluster 11: 3 members
	Cluster 12: 3 members
	Cluster 13: 3 members
	Cluster 14: 3 members
	Cluster 15: 3 members
	Cluster 16: 3 members

Name	total shared cM	Tree	Tree Person Count	Cluster
Mer	32.22	https://www.myheritage.com/pedigree-tree-5:	58	13
glen	54.53	https://www.myheritage.com/pedigree-tree-3:	106	13
Eric	37.05	https://www.myheritage.com/pedigree-tree-5'	73	13

AutoCluster on MyHeritage

DNA Company			MyHeritage	
User log onto			MyHeritage	
ID & Passwords provided to Genetic Affairs			No	
Clustering Parameters by			algorithm	
Typical min - max cM			30-400	
Results delivered by			email Zip file	
Visual Clustering			html file	
Printable Cluster List			Excel	
Link to Matches			Yes	
Link to Trees			Yes	
Input for DNA Painter			No	
Typical results: # matches			102	
Typical results: # clusters			19	

AutoCluster Comparison for 3 Vendors

DNA Company	23&Me	FTDNA	MyHeritage	
User log onto	Genetic Affairs	Genetic Affairs	MyHeritage	
ID & Passwords provided to GA	Yes	Yes	No	
Clustering Parameters by	user	user	algorithm	
Typical min - max cM	50-250 default	50-250 default	30-400	
Results delivered by	email Zip file	email Zip file	email Zip file	
Visual Clustering	html file	html file	html file	
Printable Cluster List	Excel	Excel	Excel	
Link to Matches	Yes	No	Yes	
Link to Trees	No	Yes	Yes	
Input for DNA Painter	Yes	Yes	No	
Typical results: # matches	29	134	102	
Typical results: # clusters	7	24	19	

AutoCluster Comparison
for 4 Vendors

DNA Company	23&Me	FTDNA	MyHeritage	GEDmatch
User log onto	Genetic Affairs	Genetic Affairs	MyHeritage	GEDmatch Tier 1
ID & Passwords provided to GA	Yes	Yes	No	No
Cost per run	US$0.75	US$0.75	no charge	no charge
Clustering Parameters by	user	user	algorithm	user
Typical min - max cM	50-250 default	50-250 default	30-400	25-250 test
Results delivered by	email Zip file	email Zip file	email Zip file	online / save html
Visual Clustering	html file	html file	html file	on-line
Printable Cluster List	Excel	Excel	Excel	screenshot/AutoTree
Link to Matches	Yes	No	Yes	Multi Kit Analysis
Link to Trees	No	Yes	Yes	Yes
Input for DNA Painter	Yes	Yes	No	Yes (to come)
Typical results: # matches	29	134	102	56
Typical results: # clusters	7	24	19	13

AutoCluster Comparison
for 4 Vendors – More Options

DNA Company	23&Me	FTDNA	MyHeritage	GEDmatch
User log onto	Genetic Affairs	Genetic Affairs	MyHeritage	GEDmatch Tier 1
ID & Passwords provided to GA	Yes	Yes	No	No
Cost per run	US$0.75	US$0.75	no charge	no charge
Clustering Parameters by	user	user	algorithm	user
Typical min - max cM	50-250 default	50-250 default	30-400	25-250 test
Results delivered by	email Zip file	email Zip file	email Zip file	online / save html
Visual Clustering	html file	html file	html file	on-line
Printable Cluster List	Excel	Excel	Excel	screenshot/AutoTree
Link to Matches	Yes	No	Yes	Multi Kit Analysis
Link to Trees	No	Yes	Yes	Yes
Input for DNA Painter	Yes	Yes	No	Yes (to come)
Typical results: # matches	29	134	102	56
Typical results: # clusters	7	24	19	13
AutoTree	No	Yes	No	Yes
AutoSegment / ICW	Yes	Yes	No	No

Clustering with Manual Inputs
"Run AutoCluster with csv files"

Match Data can come from Ancestry, Living DNA, or anywhere.

Sample Match File & cM

Ver		359	maternal
Pet		278	
Lei		269	
RAl		246	
B.A		239	
Ella		102	paternal
Clif		195	
Lea		191	
Kat		184	
Sus		164	
CQ		157	
Tar		157	
J.K.		153	
Tar		152	
E.E		149	

Sample Shared Match File

Ver		Jar	
Ver		JRN	
Ver		She	
Mic		E.B	
ML		Ker	
Nar		Juli	
Nel		Dav	
Nel		terr	
Nel		ver	
Nel		j.k.	
Nel		a.w	
Llar		ver	
Llar		Kat	
Lor		Lin	
Lor		Der	

Clustering Ancestry Matches
"Ancestry DNA Match Table Maker"

Google Sheets for both Match Table and Shared Match Tables

Creator: Greg Clarke   Distributor: Family History Fanatics
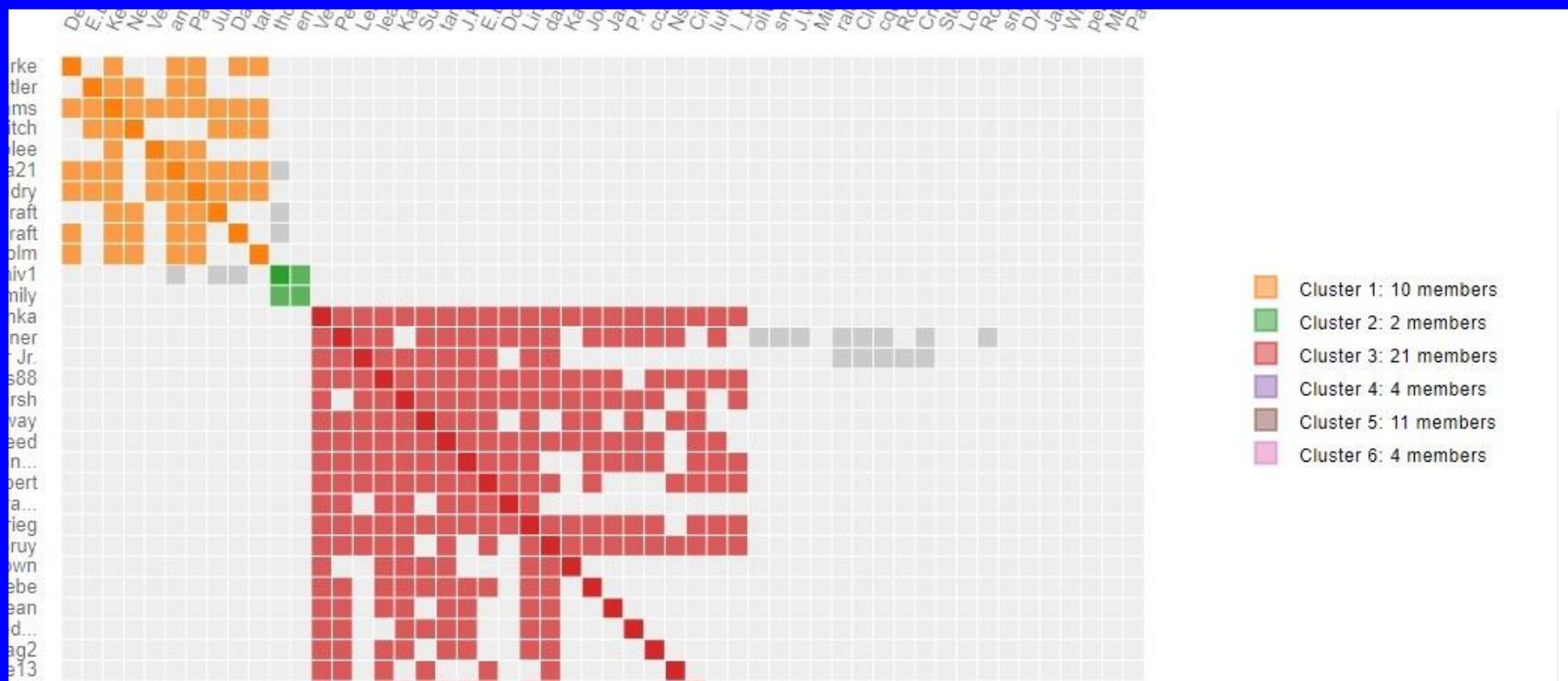
Advantages of Segment Matches vs Total cM Matches

- Less likely to be false

- Less likely to be missed

- Can match people from different vendors

- Possible to deal with pile ups

- Step towards DNA Painting.

Finding Segment Matches
Family Tree DNA (Manual Search)

* Method 1
  – Family Finder - search for person of interest – ex: E. Hunt
  – Chromo browser - ID segments – chromo 2, 175-212 Mbp
  – Family Finder   - search for shared matches with E. Hunt
  – Chromo browser - ID shared matches with same segment.
  –

* Method 2
  – Chromo browser - Download all segments (when not displaying)
  – Read segment file into Excel – find person & ID segment
  – Select correct chromosome
  – Sort matches in order of segment start position.
  – Select matching segments.

Finding Segment Matches
Family Tree DNA (Manual Display)

Creating Segment Clusters With Genetic Affairs - FTDNA

- From Family Finder Page
  - Download all matches
- From Chromosome Browser
  - Download all segments   (when not displaying)
- From Genetic Affairs
  - Run AutoSegment - for FTDNA
    » Enter file names, min-max cM, overlap cM
    » Pileup removal flag
- Segment Cluster results via email as zip files

15

Genetic Affairs - FTDNA
AutoSegment Cluster 8   11 matches

☐ Sort Display 46 clusters, too large.

Segment Cluster 8 - Chromosome 2			Total cM	Segment Million Base Pairs			Seg cM
Vendor	Match Name			Start	End	Length	
	Total Matches = 11			170.9	215.1	44.2	
FTDNA	Fred		79	170.9	215.1	44.2	37.5
FTDNA	Ernie		67	170.9	212.4	41.5	35.5
FTDNA	Eric		51	174.8	212.4	37.6	29.7
FTDNA	Robe		66	174.4	205.5	31.1	22.1
FTDNA	Park		79	174.4	204.7	30.3	21.4
FTDNA	Bran		67	174.4	204.3	29.9	20.6
FTDNA	Davi		47	172.9	195.6	22.7	18.3
FTDNA	Kevi		41	174.4	197.5	23.1	17.0
FTDNA	Dona		57	174.8	196.6	21.8	15.8
FTDNA	BARI		49	174.8	196.6	21.8	15.8
FTDNA	Kath		46	174.4	195.6	21.2	15.5

AutoSegment Cluster Comparison

DNA Company	23&Me	FTDNA	MyHeritage	GEDmatch
User log onto	23&Me	FTDNA	MyHeritage	GEDmatch T1
Cost per run	US$0.75	US$0.75	US$0.75	US$0.75
Download/Export Match File	aggregate data	Matches: CSV	Entire Matches	Run Segment Srch
Download/Export Segment File	(combined)	All Segments	shared segment	Run Triangulation
User log onto	Genetic Affairs	Genetic Affairs	Genetic Affairs	Genetic Affairs
Clustering Parameters by	user	user	user	user
Typical min - max cM	25-250 default	25-250 default	25-250 default	15-250 default
Typical segment overlap cM	15 default	15 default	15 default	9 default
Printable Cluster List	Excel	Excel	Excel	Excel
Link to Matches	Yes	email	Yes	email
Link to Trees	No	No	Yes	No
Input for DNA Painter	Yes	Yes	Yes	Yes
Typical results: # matches	262	253	195	497
Typical results: # clusters	43	46	43	92

Genetic Affairs – 4 Vendors Combined
Hybrid AutoSegment Cluster 13   47 matches

Match Cluster 109 - Segment Cluster 13 - Chromo 2		Total cM	Segment Milliion Base Pairs			
Vendor	Match Name		Start	End	Length	Seg cM
	Total matches = 47		170.4	223.3	52.9	
MyHeritage	Amy	57	174.3	223.3	49.0	45.7
23andMe	Teel	54	174.5	223.2	48.7	45.2
FTDNA	Fred	79	171.2	215.4	44.2	38.2
23andMe	Udc	37	171.0	213.8	42.8	38.1
23andMe	Rod	36	170.4	212.9	42.5	37.1
FTDNA	Ernie	67	171.2	212.7	41.5	35.9
GEDmatch	A417	34	171.0	213.8	42.8	35.4
GEDmatch	A812	33	171.0	213.3	42.3	34.6
MyHeritage	glen	54	172.4	212.1	39.7	32.8
GEDmatch	A240	30	172.4	211.9	39.5	31.6
23andMe	Hun	36	175.1	212.9	37.8	30.3
MyHeritage	Eric	37	175.0	212.9	37.9	30.1
FTDNA	Eric	51	175.1	212.7	37.6	29.8
MyHeritage	Mer	32	174.3	206.7	32.4	25.2
23andMe	Mal	23	172.7	203.0	30.3	24.2
23andMe	Wal	22	174.2	205.7	31.5	23.9
MyHeritage	Mich	29	174.3	205.8	31.5	23.2

DNA Pileups
Segments with "excess" matches

- Pileups are segments that are so common that they cannot identify recent ancestors.

- Pileups are used by DNA Painter as a warning.

- Genetic Affairs has an option to eliminate matches that overlap a pileup by 3 cM

- There is a 5 cM pileup on Chromosome 2 that would eliminate the entire Hunt cluster.

- A different logic can be used so a 5 cM pileup does not eliminate a 30 cM match.

Pileups Reported by Hong Li
Used by Genetic Affairs & DNA Painter

Chromosome	Starting position	Ending position	Genetic length (in cM)	Ratio of observed to expected IBD		
				GERMLINE Europe	GERMLINE Asia	ISCA Europe
chr9	38,293,483	72,605,261	8.15	39	13	10
chr8	10,428,647	13,469,693	7.96	38	26	2
chr21	16,344,186	19,375,168	6.91	22	15	2
chr10	44,555,093	53,240,188	7.58	22	21	2
chr22	16,051,881	25,095,451	20.82	22	22	3
chr2	85,304,243	99,558,013	6.53	21	21	2
chr1	118,434,520	153,401,108	9.95	19	33	47
chr15	20,060,673	25,145,260	10.46	15	20	42
chr17	77,186,666	78,417,478	5.66	11	7	0.1
chr15	27,115,823	30,295,750	9.29	9	3	3
chr17	59,518,083	64,970,531	6.23	9	4	4
chr2	132,695,025	141,442,636	9.16	7	0	4
chr16	19,393,068	24,031,556	6.18	6	2	5
chr2	192,352,906	198,110,229	5.04	4	2	4
Total	14 regions		119.92			

Match Density Plot
Match Length vs Pileup Length

Family Tree DNA
Match Density Chromosome 2
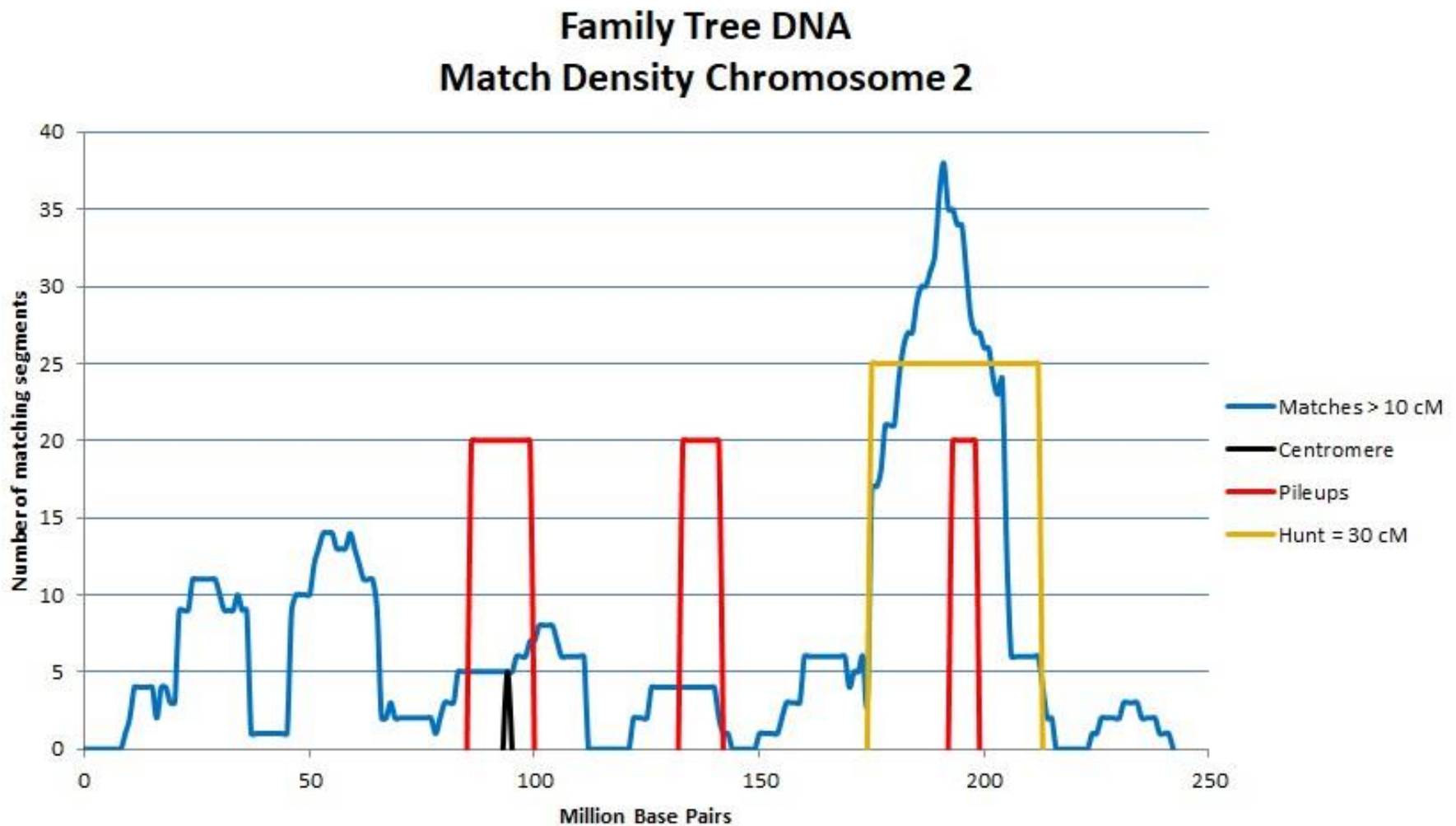
Match Density Plot
Matches including 1 cM segments

Family Tree DNA
Match Density Chromosome 2 using 1 cM segments

Legend:
- Matches > 1 cM
- Matches > 10 cM
- Centromere
- Pileups
- Hunt = 30 cM

Y-axis: Number of matching segments (0, 200, 400, 600, 800, 1000, 1200)

X-axis: Million Base Pairs (0, 50, 100, 150, 200, 250)
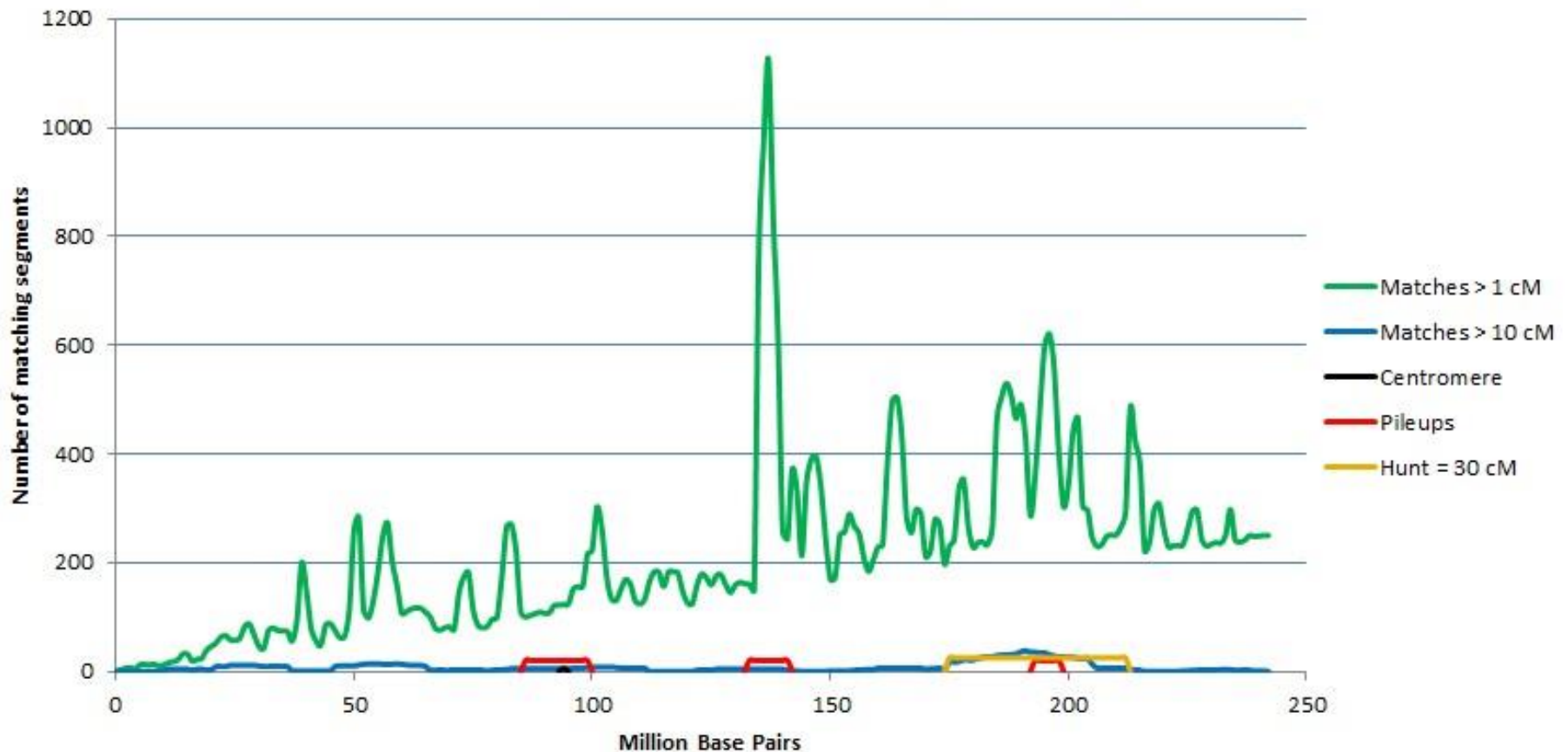
Match Density Plot
Matches minimum 20 cM

Family Tree DNA
Match Density Chromosome 2 using 20 cM segments

Legend:
- Matches > 20 cM
- Matches > 10 cM
- Centromere
- Pileups
- Hunt = 30 cM

Y-axis: Number of matching segments (0–40)
X-axis: Million Base Pairs (0–250)
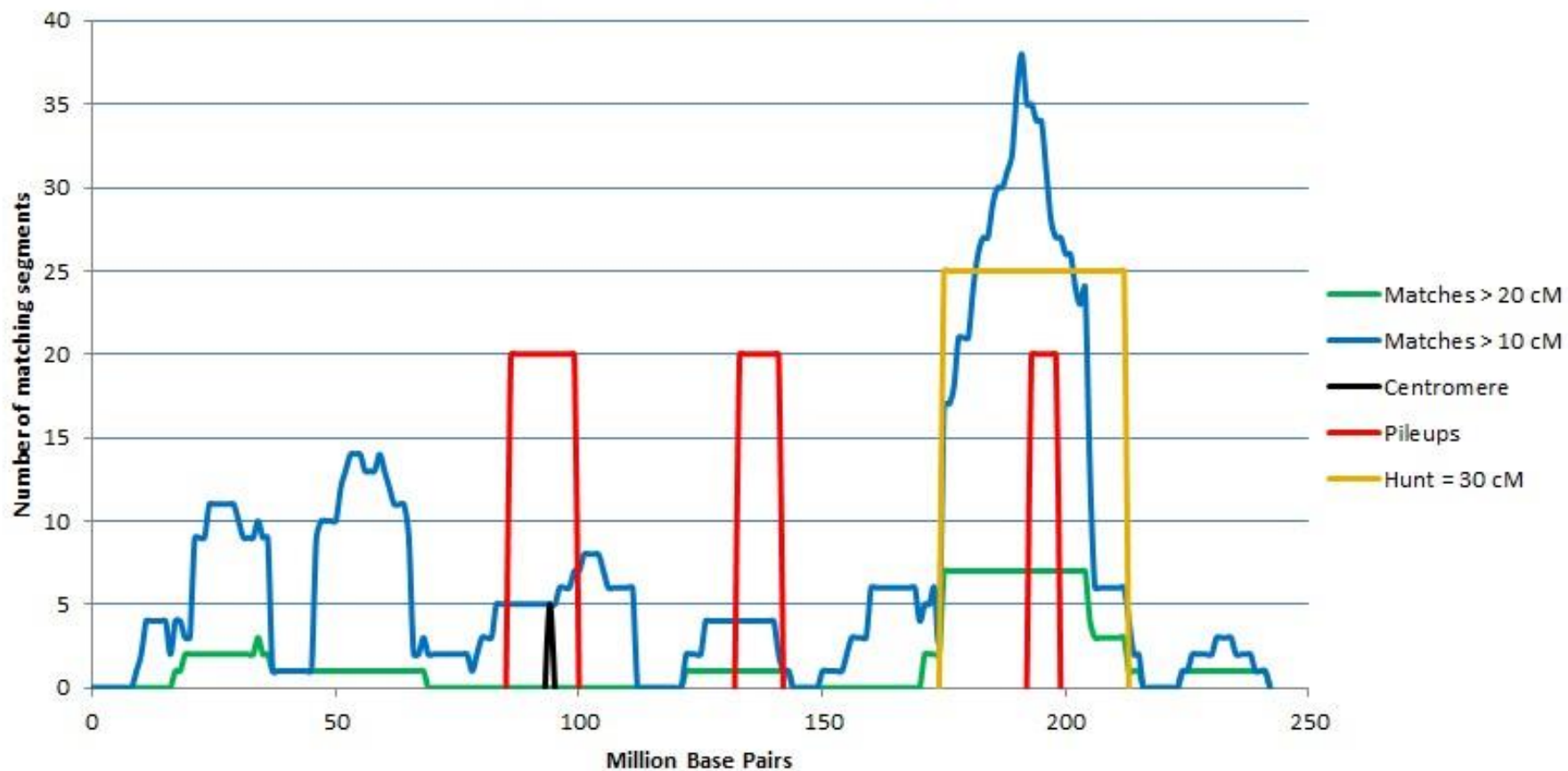
Segment Cluster after Pileup Removal
Net useful segment = total segment – pileup

Match Cluster 109 - Segment Cluster 13 - Chromo 2			Segment Milliion Base Pairs				Adjusted for Pileups	
Vendor	Match Name	Total cM	Start	End	Length	Seg cM	Length	Seg cM
	Total matches = 47		170.4	223.3	52.9			
	Pile Up Segment		192.4	198.1	5.7	5.0		
MyHeritage	Amy	57	174.3	223.3	49.0	45.7	43.3	40.7
23andMe	Tee	54	174.5	223.2	48.7	45.2	43.0	40.2
FTDNA	Fred	79	171.2	215.4	44.2	38.2	38.5	33.2
23andMe	Udc	37	171.0	213.8	42.8	38.1	37.1	33.1
23andMe	Roc	36	170.4	212.9	42.5	37.1	36.8	32.1
FTDNA	Erni	67	171.2	212.7	41.5	35.9	35.8	30.9
GEDmatch	A41	34	171.0	213.8	42.8	35.4	37.1	30.4
GEDmatch	A81	33	171.0	213.3	42.3	34.6	36.6	29.6
MyHeritage	glen	54	172.4	212.1	39.7	32.8	34.0	27.8
GEDmatch	A24	30	172.4	211.9	39.5	31.6	33.8	26.6
23andMe	Hur	36	175.1	212.9	37.8	30.3	32.1	25.3
MyHeritage	Eric	37	175.0	212.9	37.9	30.1	32.2	25.1
FTDNA	Eric	51	175.1	212.7	37.6	29.8	31.9	24.8
MyHeritage	Mer	32	174.3	206.7	32.4	25.2	26.7	20.2
23andMe	Mal	23	172.7	203.0	30.3	24.2	24.6	19.2
23andMe	Wa	22	174.2	205.7	31.5	23.9	25.8	18.9
MyHeritage	Mic	29	174.3	205.8	31.5	23.2	25.8	18.2

Segment Matches
Paternal & Maternal Combined
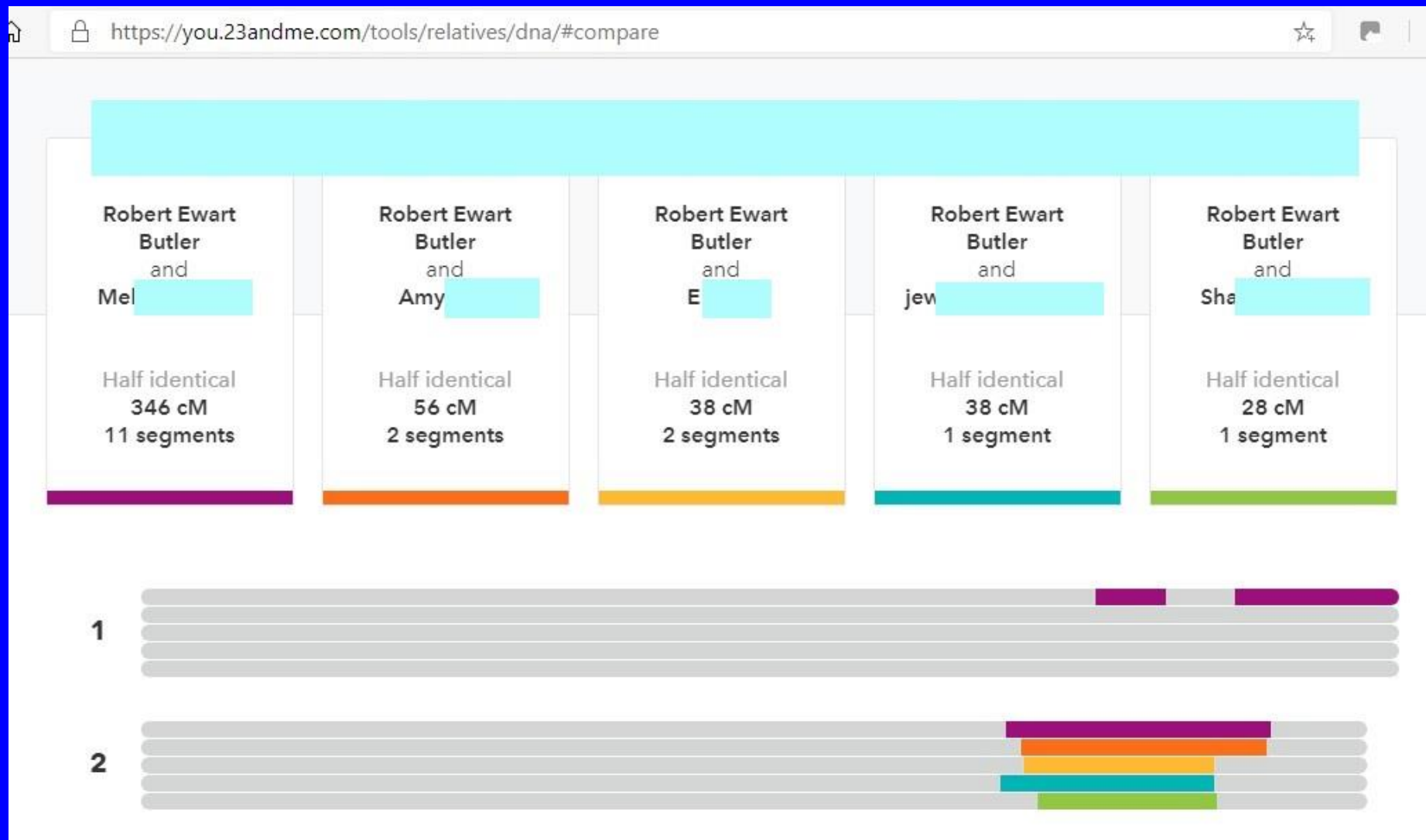
Triangulated Matches
Paternal & Maternal Separated (23&Me)

You and Amy may have Relatives in Common
Finding common relatives can help you piece together your family story.

Relative in common	You	Amy	Shared DNA
Mel	1st Cousin, Twice Removed (4.64%)	4th Cousin (0.63%)	Yes
Step	3rd Cousin (1.71%)	4th Cousin (0.33%)	No
Tho	3rd Cousin (1.48%)	5th Cousin (0.09%)	No
Chri	3rd Cousin (1.46%)	4th Cousin (0.44%)	No

You and E may have Relatives in Common
Finding common relatives can help you piece together your family story.

| Relative in common	You	E	Shared DNA
J S	4th Cousin (0.42%)	4th Cousin (0.36%)	Share to see
Geo	4th Cousin (0.39%)	5th Cousin (0.10%)	No
Rene	4th Cousin (0.38%)	Distant Cousin (0.09%)	No
Shav	4th Cousin (0.38%)	4th Cousin (0.72%)	Yes

Segment Cluster after Pileup Removal
and after Triangulation Check

Match Cluster 109 - Segment Cluster 13 - Chromo 2			Segment Milliion Base Pairs				Adjusted for Pileups	
Vendor	Match Name	Total cM	Start	End	Length	Seg cM	Length	Seg cM
Paternal TG	Total matches = 47		170.4	223.3	52.9			
Maternal TG	Pile Up Segment		192.4	198.1	5.7	5.0		
MyHeritage	Amy	57	174.3	223.3	49.0	45.7	43.3	40.7
23andMe	Tee	54	174.5	223.2	48.7	45.2	43.0	40.2
FTDNA	Fred	79	171.2	215.4	44.2	38.2	38.5	33.2
23andMe	Udo	37	171.0	213.8	42.8	38.1	37.1	33.1
23andMe	Rod	36	170.4	212.9	42.5	37.1	36.8	32.1
FTDNA	Ernie	67	171.2	212.7	41.5	35.9	35.8	30.9
GEDmatch	A41	34	171.0	213.8	42.8	35.4	37.1	30.4
GEDmatch	A81	33	171.0	213.3	42.3	34.6	36.6	29.6
MyHeritage	glen	54	172.4	212.1	39.7	32.8	34.0	27.8
GEDmatch	A24(	30	172.4	211.9	39.5	31.6	33.8	26.6
23andMe	Hun	36	175.1	212.9	37.8	30.3	32.1	25.3
MyHeritage	Eric	37	175.0	212.9	37.9	30.1	32.2	25.1
FTDNA	Eric	51	175.1	212.7	37.6	29.8	31.9	24.8
MyHeritage	Mer	32	174.3	206.7	32.4	25.2	26.7	20.2
23andMe	Mal	23	172.7	203.0	30.3	24.2	24.6	19.2
23andMe	Wal	22	174.2	205.7	31.5	23.9	25.8	18.9
MyHeritage	Mich	29	174.3	205.8	31.5	23.2	25.8	18.2

Total cM Cluster of Ancestry Matches
Manual Group Cluster & Google Sheets Formatting

Highlighted names also on segment clusters.

Name/Manager	Total cM	People in Tree	Notes	
B.eile	42 cM	Public 12		
Julia	40 cM	Private 56		
Glori	37 cM	No Trees		
Dian	33 cM	Public 47	Common ancestor	John Wilson & Jane Hunt
Dave	31 cM	Unlinked		
judy9	29 cM	Public 65		
Kath	28 cM	Public 70	Common ancestor	Frances Harper ??
Shar	28 cM	Private 11	Common ancestor	Frances Harper ??
joan	26 cM	Private 1,229	Common ancestor	John Wilson & Jane Hunt
Dani	26 cM			
hunt	25 cM	Private 3,656		
jrske	24 cM	Public 167		
Dore	24 cM	Public 633		
Mary	23 cM	Unlinked		
Amb	22 cM	Public 2,340	Common ancestor	Frances Harper ??
betty	22 cM	Public 116		
Fred	21 cM	Public 2	Common ancestor	Frances Harper ??

28
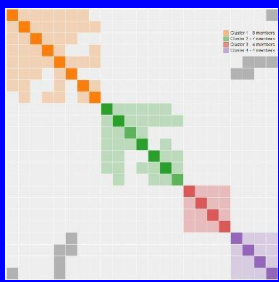
Genetic Affairs
Main Programs

Websites

Show websites
Show registered FamilyTreeDNA and 23andme websites

Register a new website
Register a new FamilyTreeDNA or 23andme website using your login credentials

AutoCluster analyses

Run AutoCluster
Run AutoCluster analysis for FamilyTreeDNA or 23andme profiles. Select the website and profile to start the AutoCluster analysis.

Run AutoTree
Run AutoTree analysis for FamilyTreeDNA profiles. Select the FTDNA profile to start the AutoTree analysis.

AutoSegment analyses

Run AutoSegment
Run AutoSegment analysis using locally downloaded DNA segments from MyHeritage, FTDNA, 23andme or GEDmatch

Run hybrid AutoSegment
Run a hybrid AutoSegment analysis and combine DNA segments from MyHeritage, FTDNA, 23andme or GEDmatch profiles

Genetic Affairs Summary
Primary Functionality - Sorting Matches

- Clustering based on Total cM or Segment matches
- Costs

DNA Company		23&Me	FTDNA	MyHeritage	GEDmatch	Ancestry CSV
AutoCluster,	$/run	US$0.75	US$0.75	0	0	US$0.50
AutoSegment,	$/run	US$0.75	US$0.75	US$0.75	US$0.75	-
Hybrid AutoSegment for # of vendors:			2	3	4	-
	$/run		US$1.00	US$1.25	US$1.50	-

- All segment programs provide input to DNA Painter.
- Present pileup removal logic not recommended
- Triangulation needed for segment matches
    - Some development is underway

Clusters & Segments

 Questions ?