

# Supplementary Material for “Identifying Mutation Regions for Closely Related Individuals without a Known Pedigree”

## **1 The experiments for pedigrees containing 5 generations**

We study different sets of input individuals in the latest two generations of Pedigree 2-4 in the paper, where there are 5 generations in those pedigrees. Those different sets of input individuals in the latest two generations in the pedigree are given in Figures 1-3. The results are shown in Table 1-3. We can see that the values of recall are very close to 100% in all the cases. The value of precision is getting better when the number of diseased individuals increases.

## **2 The experiments for pedigrees containing 6 generations**

We also do experiments on the pedigrees containing 6 generations. Here we consider Pedigree 5-8 as shown in Figure 4-7. The results are shown as Table 3 in the paper.

We study different sets of input individuals in the latest two generations of Pedigree 5-8, where there are 6 generations in those pedigrees. Those different sets of input individuals in the latest two generations in the pedigree are given in Figures 8-11. The results are shown in Table 4-7. We can see that the values of recall are more than 90% in most cases. The value of precision is getting better when the number of diseased individuals increases. The behavior is similar to that of 5 generations.

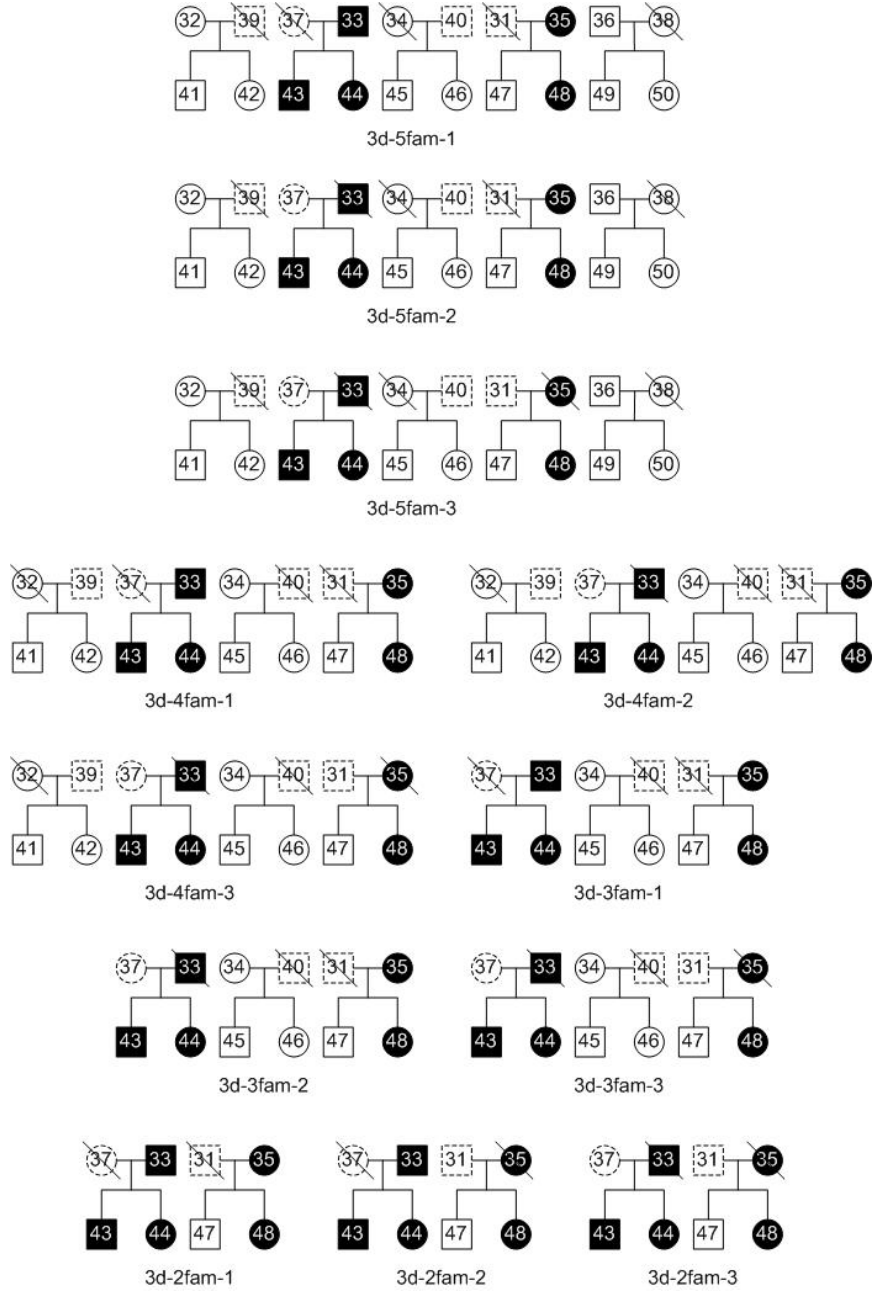


Figure 1: The different sets of input individuals based on Pedigree 2 in the paper.

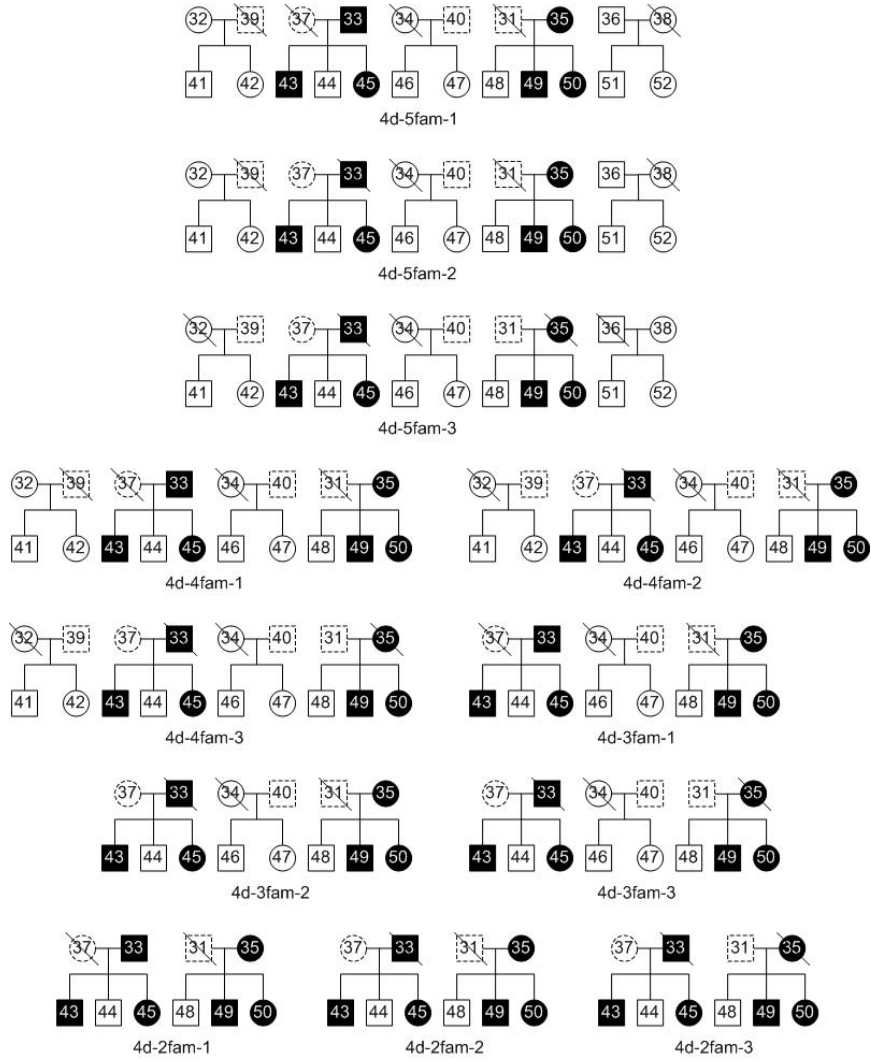


Figure 2: The different sets of input individuals based on Pedigree 3 in the paper.

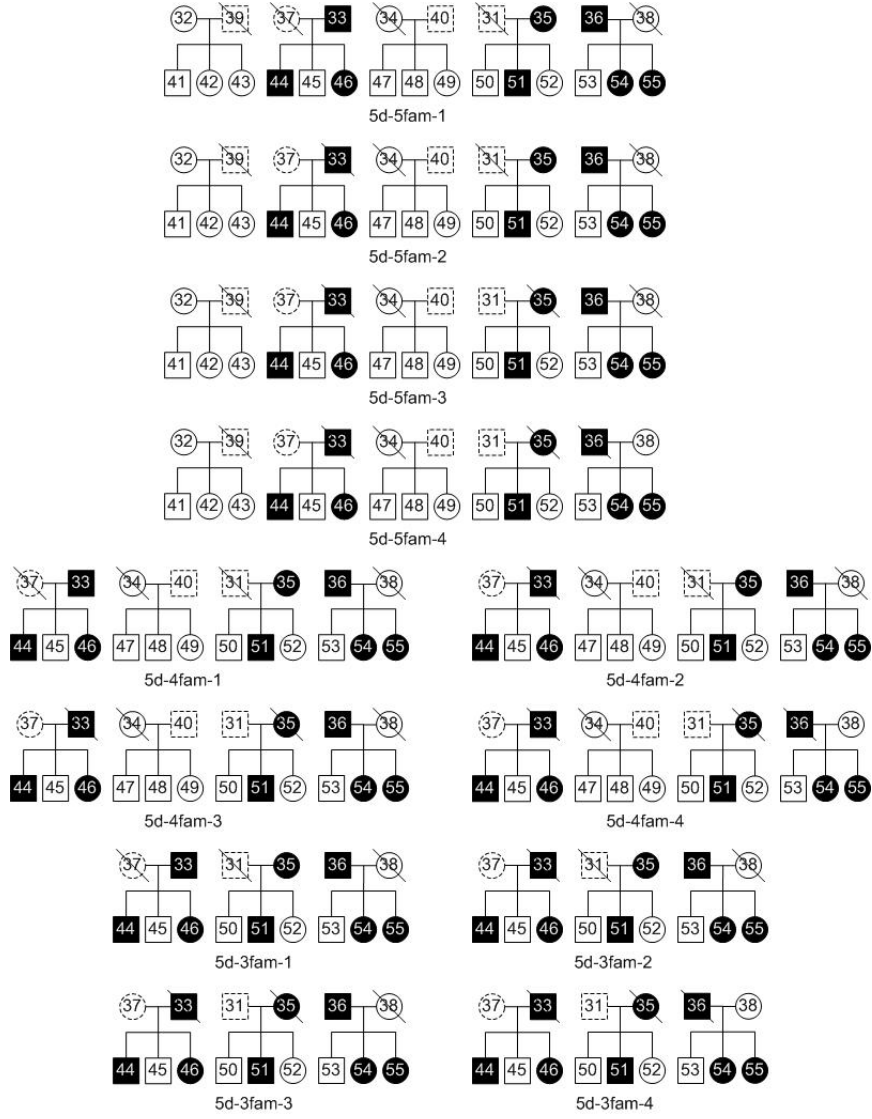


Figure 3: The different sets of input individuals based on Pedigree 4 in the paper.

input	precision	recall
3d-5fam-1	79.83%	99.04%
3d-5fam-2	74.71%	98.54%
3d-5fam-3	65.29%	97.73%
3d-4fam-1	78.80%	99.27%
3d-4fam-2	68.07%	98.63%
3d-4fam-3	59.62%	95.75%
3d-3fam-1	76.77%	99.26%
3d-3fam-2	66.28%	98.66%
3d-3fam-3	56.29%	95.06%
3d-2fam-1	72.23%	97.99%
3d-2fam-2	57.69%	95.53%
3d-2fam-3	54.67%	94.32%

Table 1: Results on Figure 1

input	precision	recall
4d-5fam-1	87.44%	96.53%
4d-5fam-2	85.13%	97.11%
4d-5fam-3	80.54%	97.41%
4d-4fam-1	85.85%	96.70%
4d-4fam-2	82.04%	97.08%
4d-4fam-3	78.82%	97.68%
4d-3fam-1	83.81%	96.49%
4d-3fam-2	80.54%	96.58%
4d-3fam-3	75.52%	97.58%
4d-2fam-1	83.78%	96.37%
4d-2fam-2	79.76%	96.76%
4d-2fam-3	74.53%	97.45%

Table 2: Results on Figure 2

input	precision	recall
5d-5fam-1	91.35%	99.26%
5d-5fam-2	90.79%	99.12%
5d-5fam-3	90.55%	99.14%
5d-5fam-4	88.67%	99.08%
5d-4fam-1	91.02%	99.40%
5d-4fam-2	90.43%	99.22%
5d-4fam-3	90.56%	99.21%
5d-4fam-4	88.76%	99.26%
5d-3fam-1	92.26%	99.42%
5d-3fam-2	92.25%	99.18%
5d-3fam-3	92.16%	98.97%
5d-3fam-4	89.84%	99.19%

Table 3: Results on Figure 3

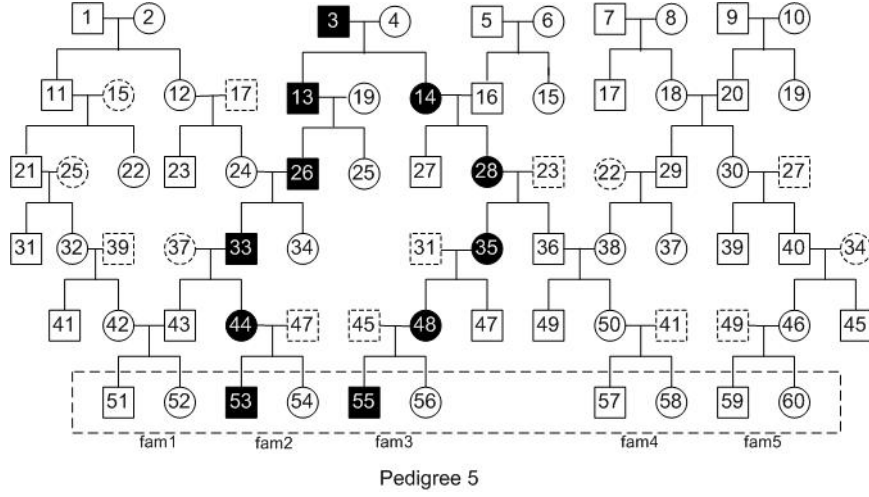
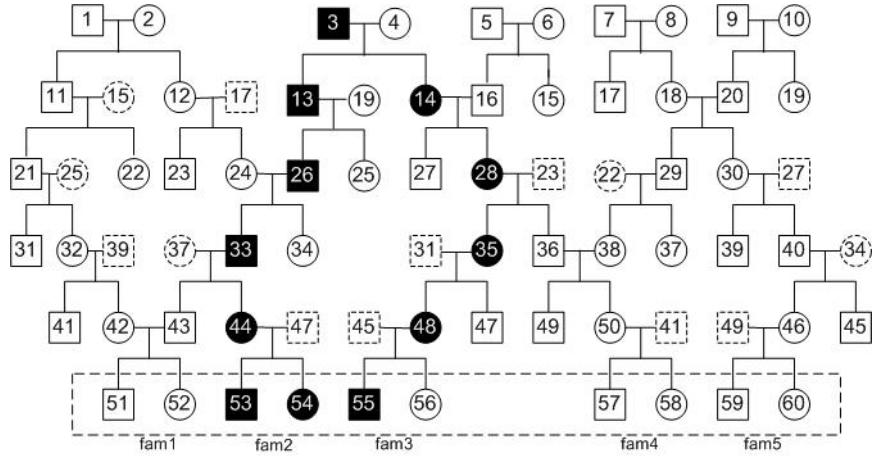
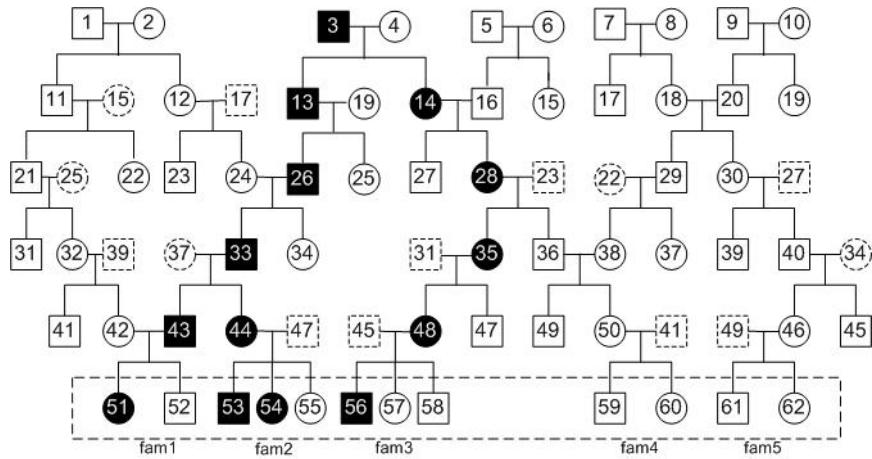


Figure 4: Pedigree 5: a pedigree containing 6 generations with 2 diseased individuals in the input.



Pedigree 6

Figure 5: Pedigree 6: a pedigree containing 6 generations with 3 diseased individuals in the input.



Pedigree 7

Figure 6: Pedigree 7: a pedigree containing 6 generations with 4 diseased individuals in the input.

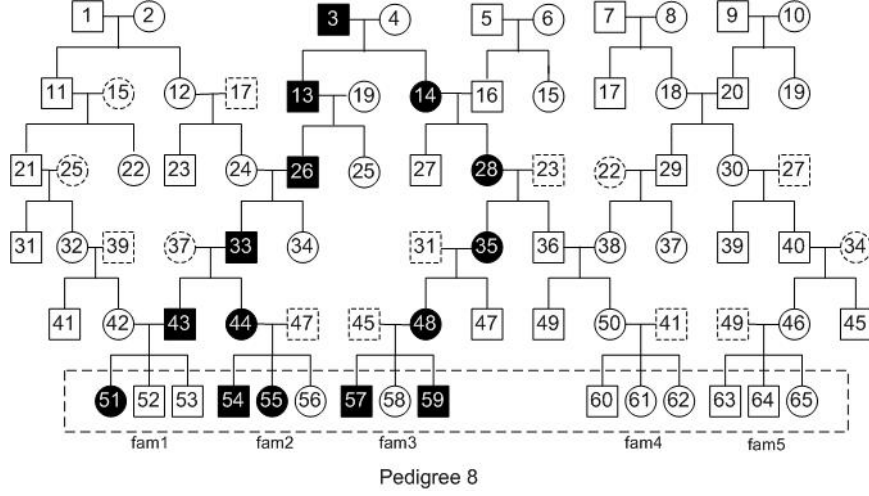


Figure 7: Pedigree 8: a pedigree containing 6 generations with 5 diseased individuals in the input.

input	precision	recall
6g-2d-5fam-1	72.38%	98.34%
6g-2d-5fam-2	61.48%	98.25%
6g-2d-5fam-3	53.47%	97.43%
6g-2d-4fam-1	69.38%	98.43%
6g-2d-4fam-2	56.54%	97.75%
6g-2d-4fam-3	49.21%	96.76%
6g-2d-3fam-1	65.74%	98.38%
6g-2d-3fam-2	47.10%	96.59%
6g-2d-3fam-3	41.66%	89.74%
6g-2d-2fam-1	61.65%	97.47%
6g-2d-2fam-2	37.10%	90.78%
6g-2d-2fam-3	34.01%	80.63%

Table 4: Results on Figure 8





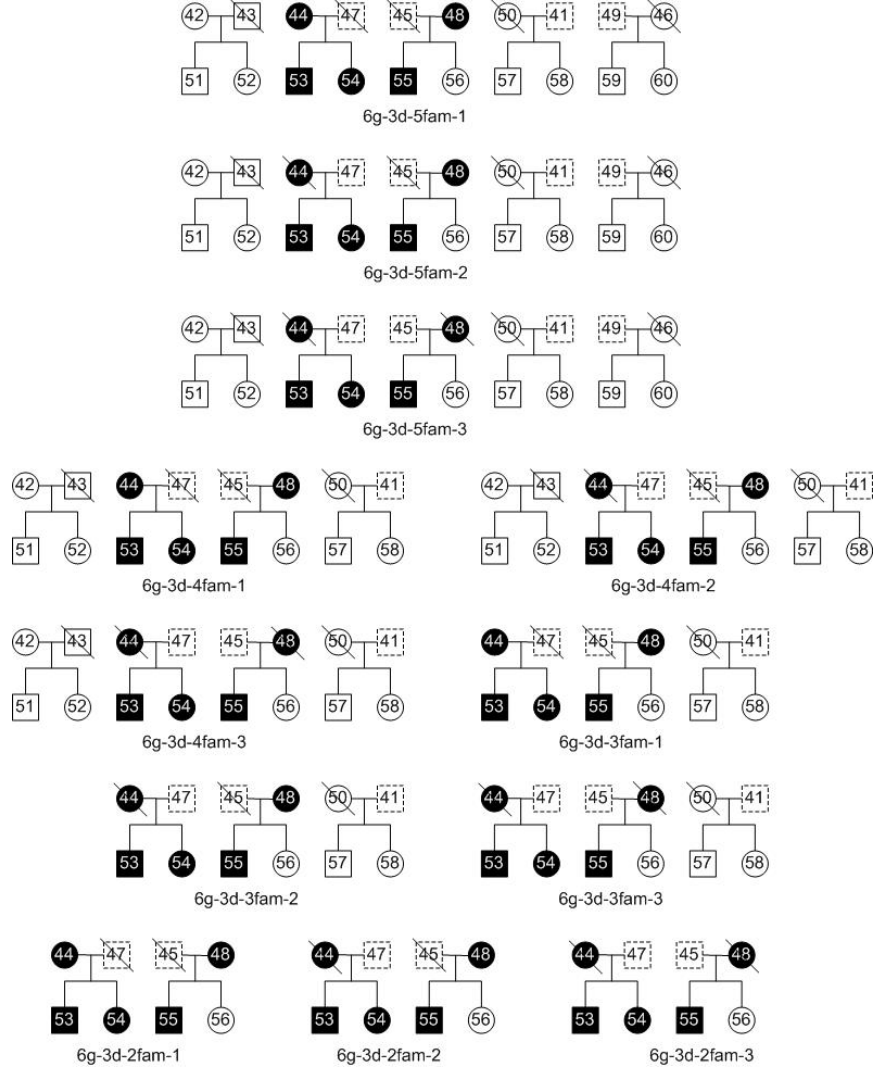


Figure 9: The different sets of input individuals based on Pedigree 6

input	precision	recall
6g-3d-5fam-1	70.18%	97.40%
6g-3d-5fam-2	65.75%	98.17%
6g-3d-5fam-3	54.38%	97.12%
6g-3d-4fam-1	66.27%	98.00%
6g-3d-4fam-2	58.36%	96.53%
6g-3d-4fam-3	49.09%	95.46%
6g-3d-3fam-1	62.53%	97.36%
6g-3d-3fam-2	50.73%	95.05%
6g-3d-3fam-3	41.44%	91.49%
6g-3d-2fam-1	60.18%	96.28%
6g-3d-2fam-2	45.59%	92.51%
6g-3d-2fam-3	35.73%	83.58%

Table 5: Results on Figure 9

input	precision	recall
6g-4d-5fam-1	90.16%	97.10%
6g-4d-5fam-2	90.04%	97.16%
6g-4d-5fam-3	89.05%	97.20%
6g-4d-5fam-4	84.16%	97.71%
6g-4d-4fam-1	90.58%	97.42%
6g-4d-4fam-2	90.14%	97.49%
6g-4d-4fam-3	88.74%	97.83%
6g-4d-4fam-4	82.41%	98.34%
6g-4d-3fam-1	90.30%	97.63%
6g-4d-3fam-2	89.41%	97.69%
6g-4d-3fam-3	87.61%	98.01%
6g-4d-3fam-4	81.54%	98.51%

Table 6: Results on Figure 10

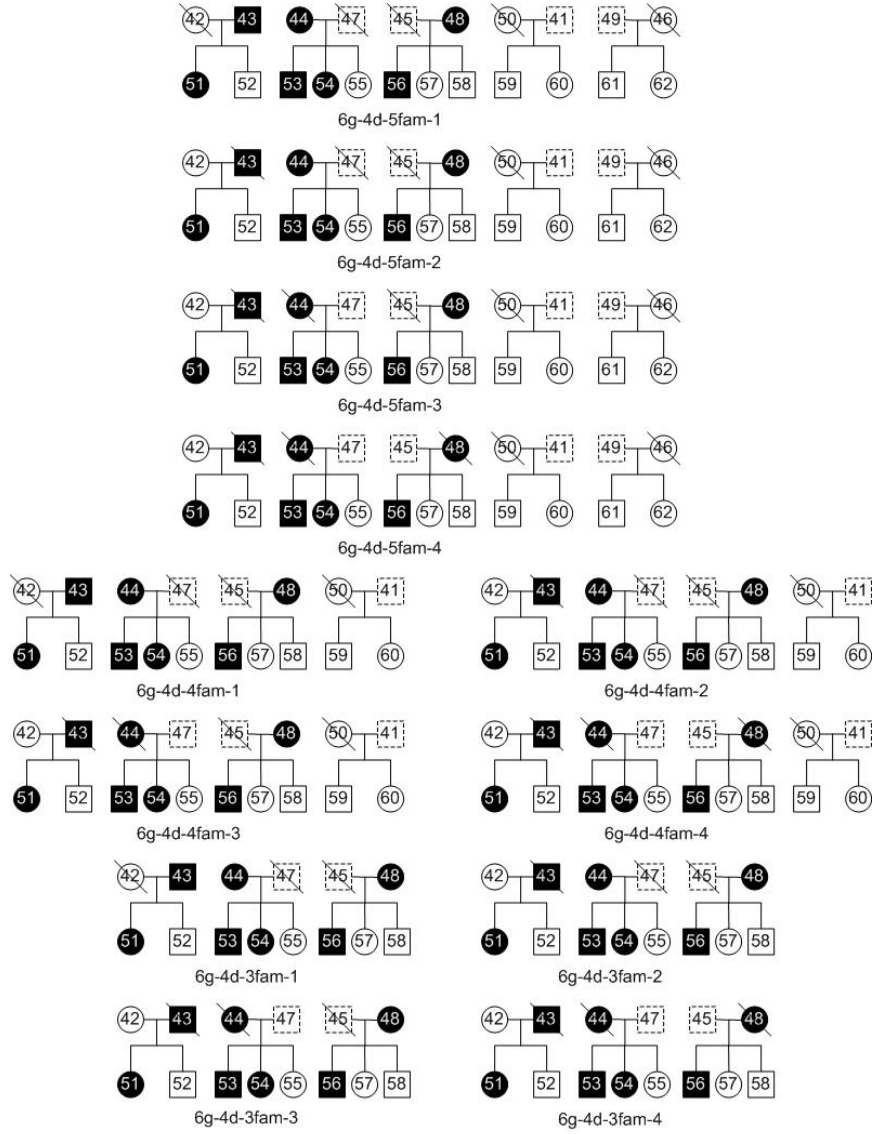


Figure 10: The different sets of input individuals based on Pedigree 7

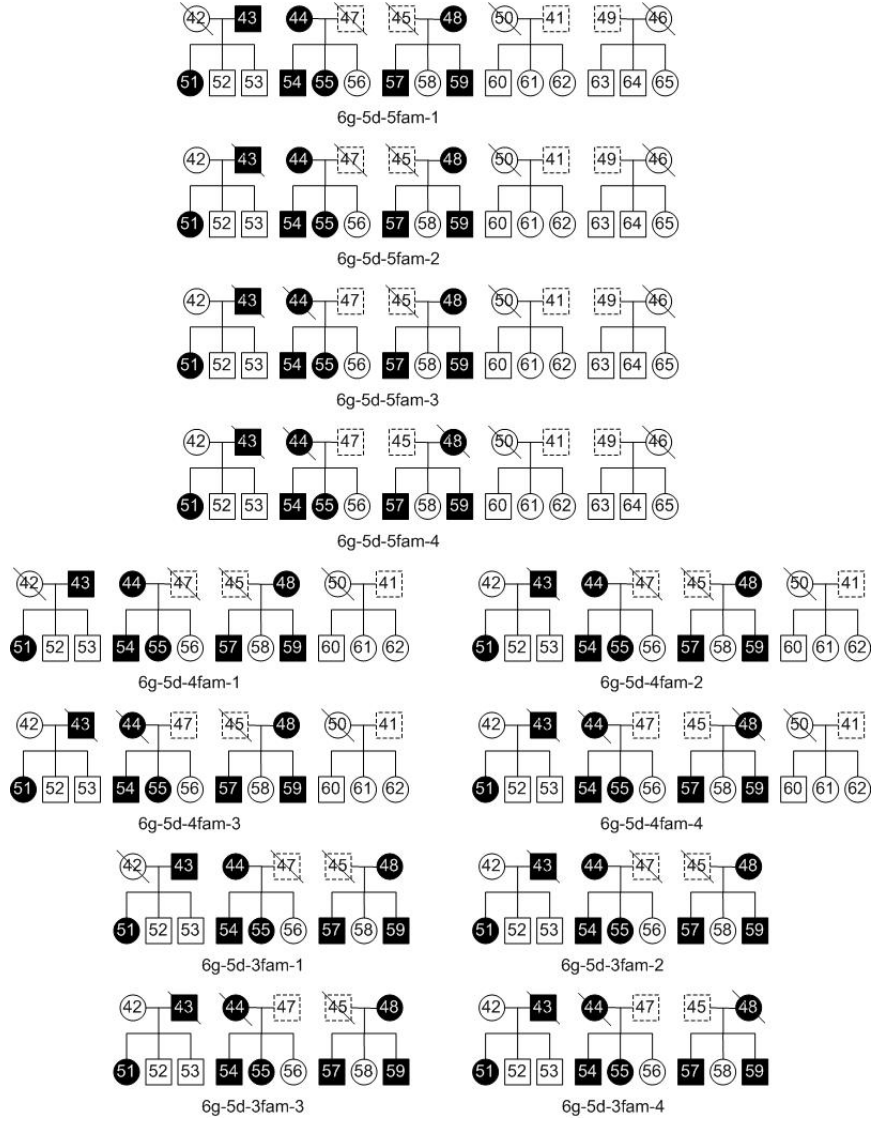


Figure 11: The different sets of input individuals based on Pedigree 8

input	precision	recall
6g-5d-5fam-1	89.56%	96.58%
6g-5d-5fam-2	88.91%	96.58%
6g-5d-5fam-3	87.72%	96.35%
6g-5d-5fam-4	87.28%	97.08%
6g-5d-4fam-1	89.56%	97.08%
6g-5d-4fam-2	89.11%	97.05%
6g-5d-4fam-3	87.69%	96.85%
6g-5d-4fam-4	86.67%	97.08%
6g-5d-3fam-1	89.42%	97.08%
6g-5d-3fam-2	89.05%	97.08%
6g-5d-3fam-3	88.04%	97.35%
6g-5d-3fam-4	86.40%	97.58%

Table 7: Results on Figure 11

### 3 The experiments for pedigrees containing 7 generations

The pedigrees shown in Figure 12-15 contain 7 generations and 2, 3, 4, 5 diseased individuals in the latest generation. Only the individuals in the latest generation are the input individuals. The experiment results are shown in Table 4 in the paper.

Figure 16-19 show the different sets of input individuals in the latest two generations of Pedigree 9-12. The results are shown in Table 8-11. The performance of our program for 7 generations is similar to that for 5 and 6 generations but slightly worse than them. We do 200 experiments for all the cases mentioned above.

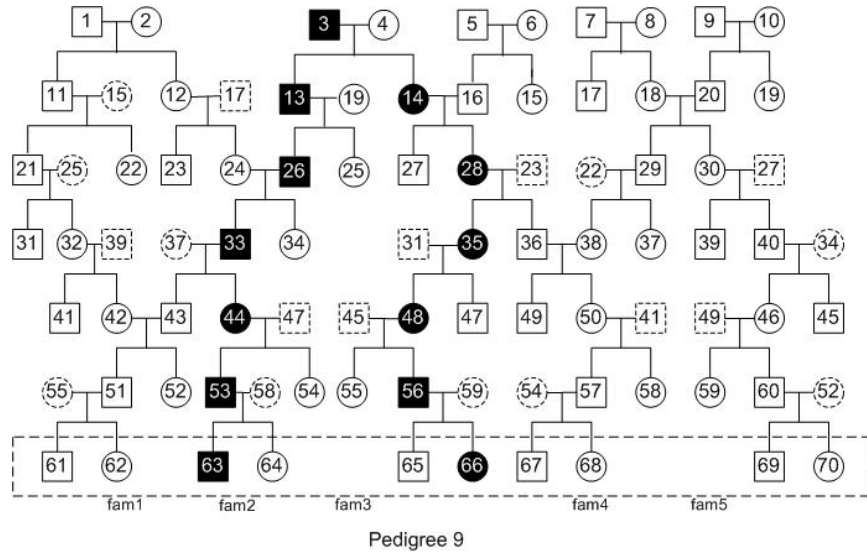


Figure 12: Pedigree 9: a pedigree containing 7 generations with 2 diseased individuals in the input.

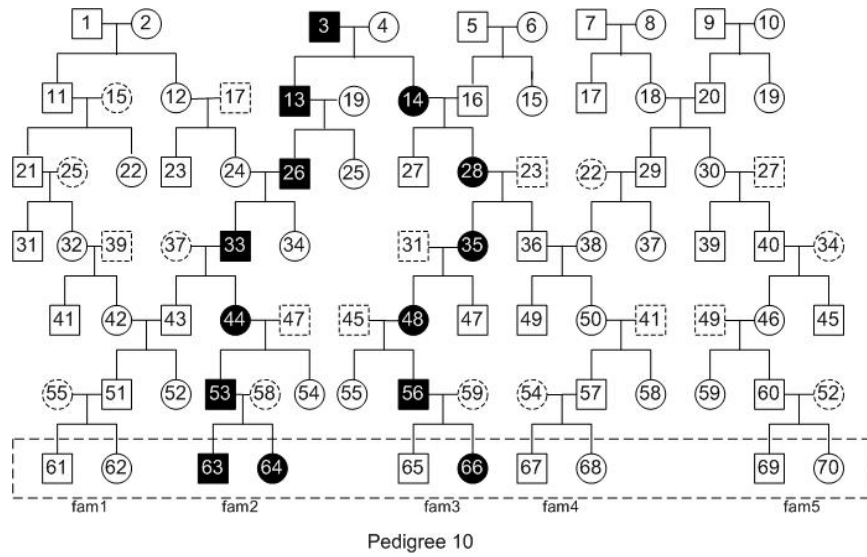


Figure 13: Pedigree 10: a pedigree containing 7 generations with 3 diseased individuals in the input.

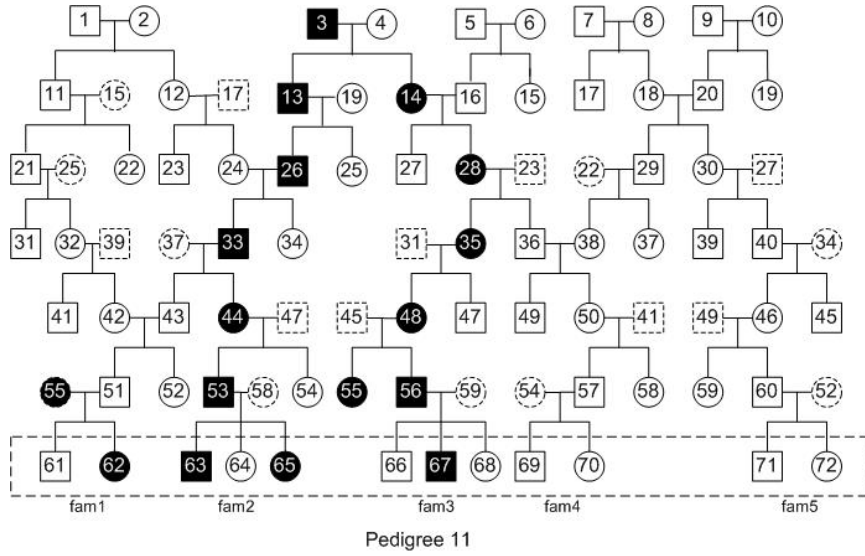


Figure 14: Pedigree 11: a pedigree containing 7 generations with 4 diseased individuals in the input.

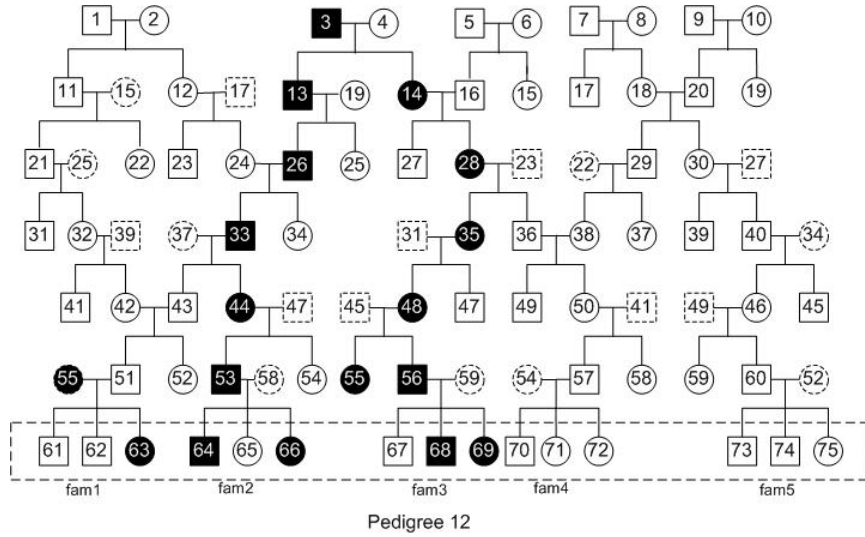


Figure 15: Pedigree 12: a pedigree containing 7 generations with 5 diseased individuals in the input.



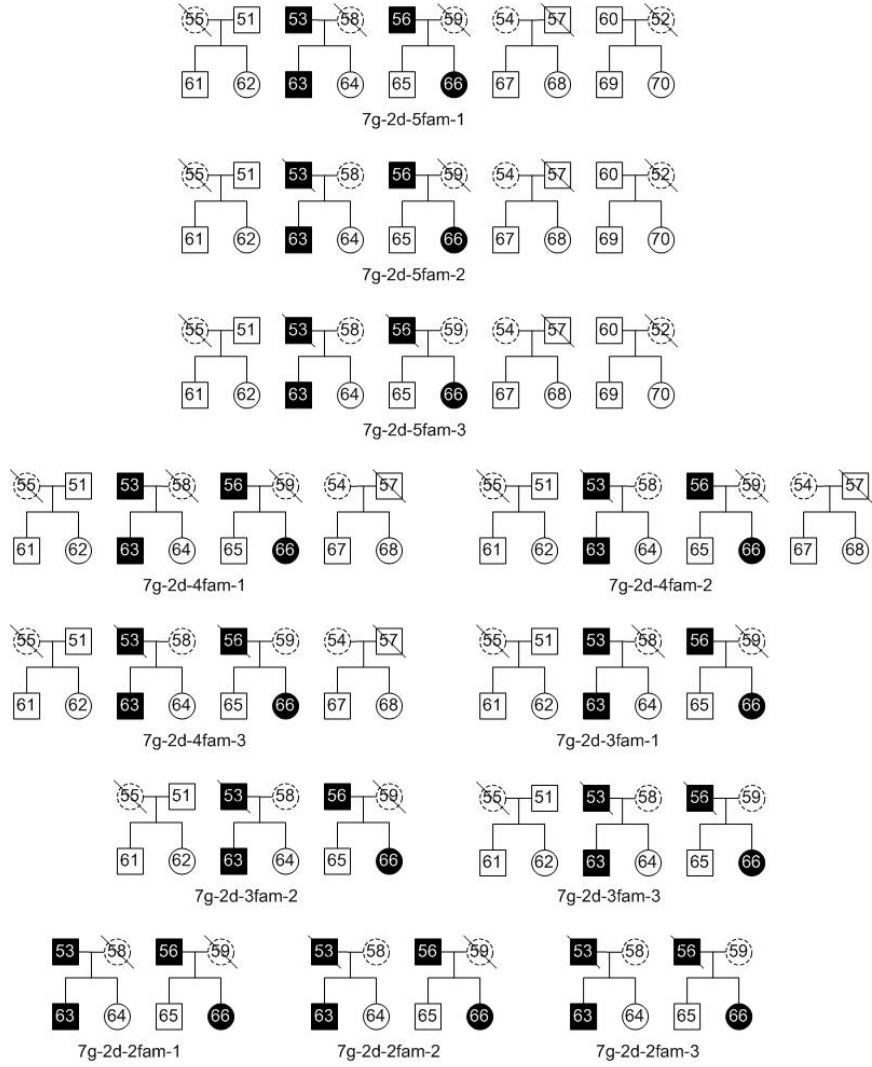


Figure 16: The different sets of input individuals based on Pedigree 9

input	precision	recall
7g-2d-5fam-1	74.83%	96.79%
7g-2d-5fam-2	64.95%	96.80%
7g-2d-5fam-3	55.08%	97.41%
7g-2d-4fam-1	69.49%	96.87%
7g-2d-4fam-2	58.55%	96.87%
7g-2d-4fam-3	50.80%	96.32%
7g-2d-3fam-1	58.11%	95.29%
7g-2d-3fam-2	51.23%	94.57%
7g-2d-3fam-3	43.30%	93.78%
7g-2d-2fam-1	48.75%	92.04%
7g-2d-2fam-2	46.30%	92.14%
7g-2d-2fam-3	37.80%	87.33%

Table 8: Results on Figure 16

input	precision	recall
7g-3d-5fam-1	77.44%	98.57%
7g-3d-5fam-2	72.88%	98.67%
7g-3d-5fam-3	61.93%	97.89%
7g-3d-4fam-1	72.40%	98.27%
7g-3d-4fam-2	66.60%	98.76%
7g-3d-4fam-3	54.46%	97.34%
7g-3d-3fam-1	57.48%	97.76%
7g-3d-3fam-2	57.71%	96.57%
7g-3d-3fam-3	45.69%	92.68%
7g-3d-2fam-1	50.98%	95.24%
7g-3d-2fam-2	55.19%	94.11%
7g-3d-2fam-3	42.25%	89.19%

Table 9: Results on Figure 17

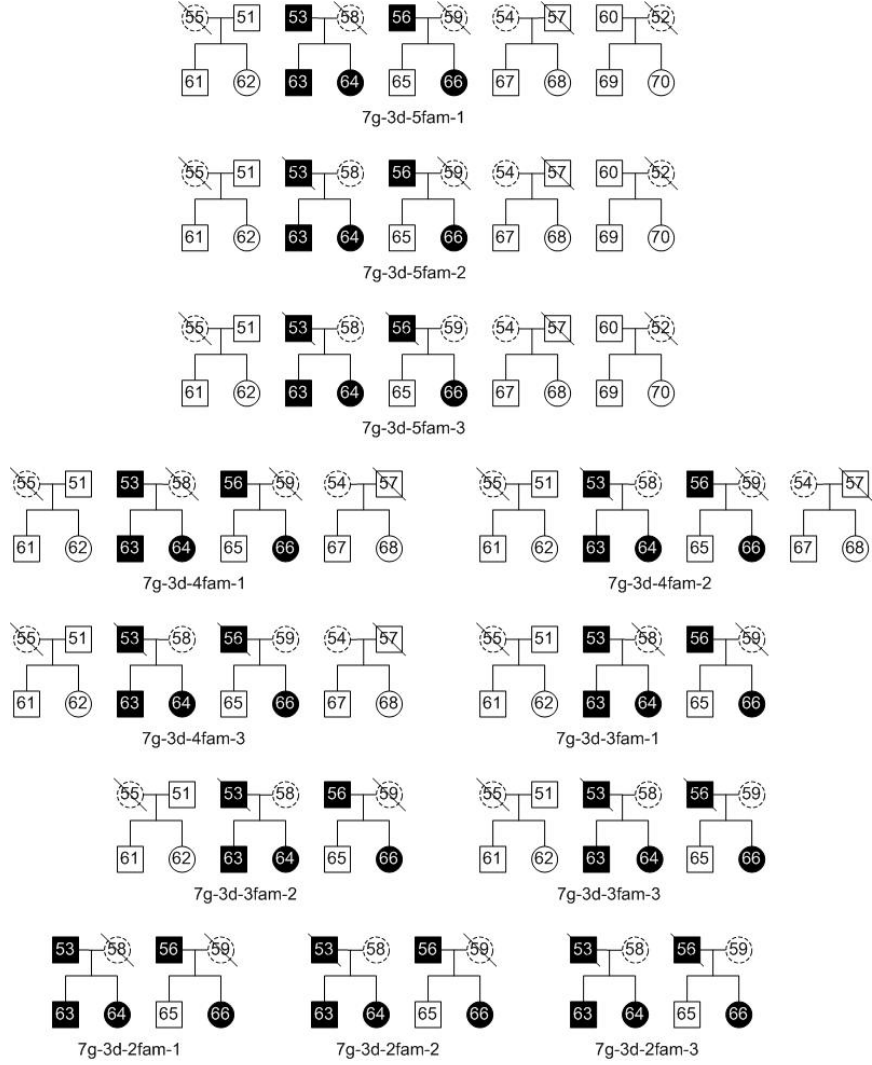


Figure 17: The different sets of input individuals based on Pedigree 10

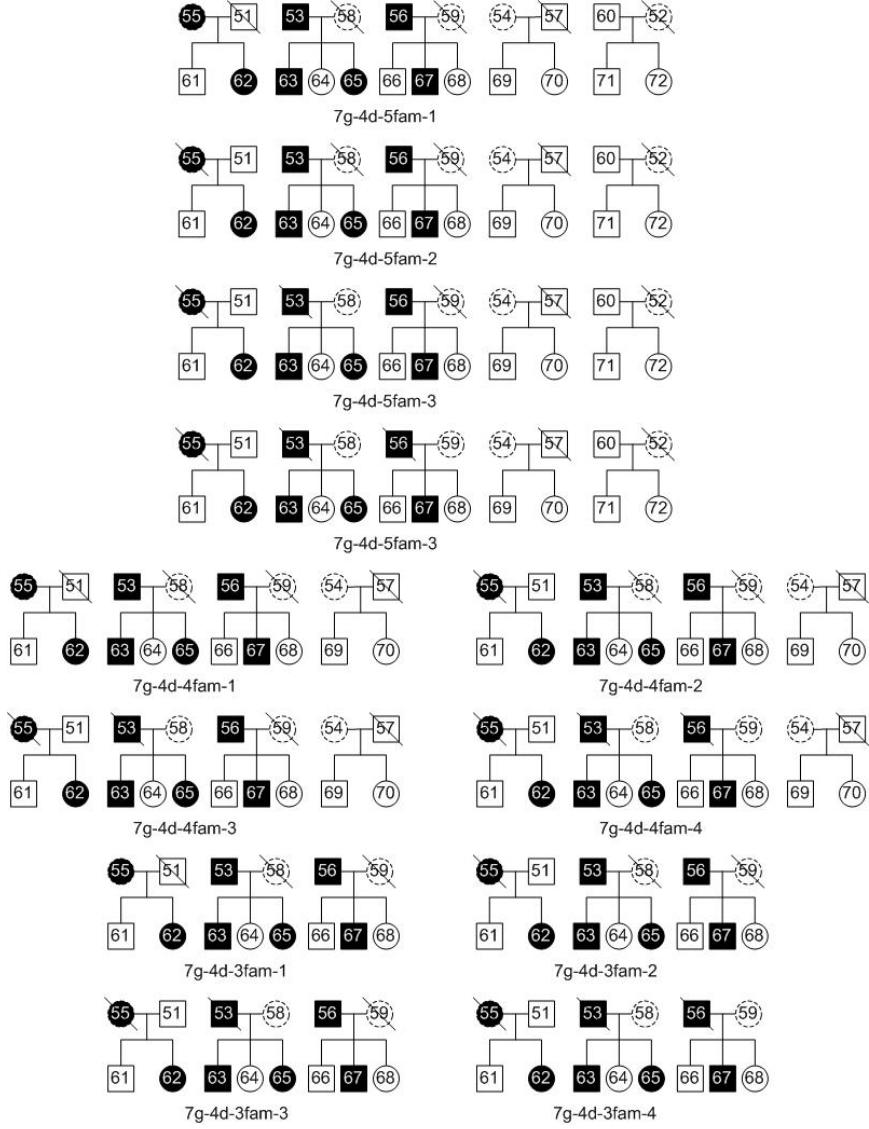


Figure 18: The different sets of input individuals based on Pedigree 11

input	precision	recall
7g-4d-5fam-1	87.51%	97.21%
7g-4d-5fam-2	88.43%	97.25%
7g-4d-5fam-3	86.93%	97.37%
7g-4d-5fam-4	85.25%	97.30%
7g-4d-4fam-1	87.53%	97.28%
7g-4d-4fam-2	88.58%	97.18%
7g-4d-4fam-3	86.63%	97.41%
7g-4d-4fam-4	83.31%	97.40%
7g-4d-3fam-1	85.66%	97.12%
7g-4d-3fam-2	86.29%	97.28%
7g-4d-3fam-3	85.34%	97.43%
7g-4d-3fam-4	81.05%	97.32%

Table 10: Results on Figure 18

input	precision	recall
7g-5d-5fam-1	91.02%	96.57%
7g-5d-5fam-2	91.20%	96.66%
7g-5d-5fam-3	89.45%	96.70%
7g-5d-5fam-4	88.32%	96.46%
7g-5d-4fam-1	90.85%	96.92%
7g-5d-4fam-2	91.22%	96.88%
7g-5d-4fam-3	88.87%	96.92%
7g-5d-4fam-4	87.45%	96.59%
7g-5d-3fam-1	89.62%	96.99%
7g-5d-3fam-2	90.33%	96.98%
7g-5d-3fam-3	89.24%	96.79%
7g-5d-3fam-4	87.51%	96.94%

Table 11: Results on Figure 19

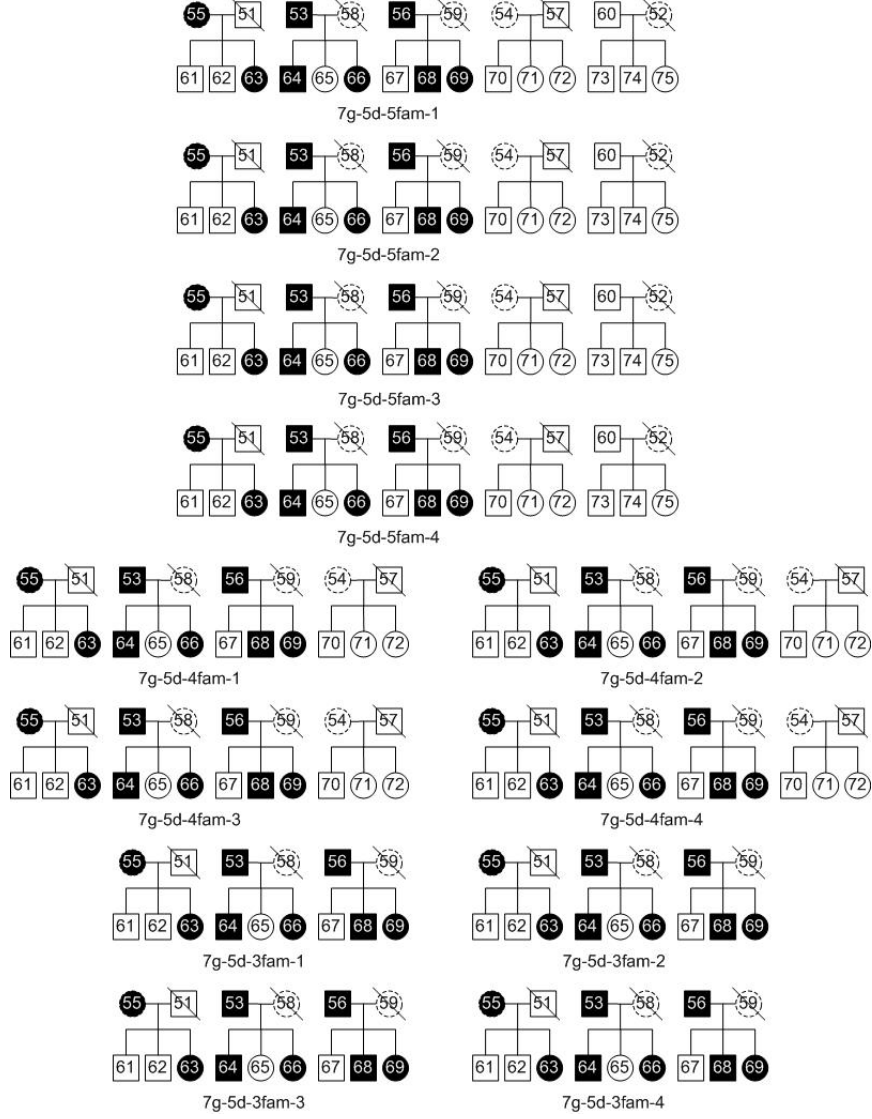


Figure 19: The different sets of input individuals based on Pedigree 12