

Statistical power in missing person cases



Lecture 2: Who should be exhumed?

Magnus Dehli Vigeland

August 25th 2021

PROYECTO DE COOPERACIÓN TRIANGULAR. ARGENTINA - COLOMBIA – UNIÓN EUROPEA.



Overview - last lecture

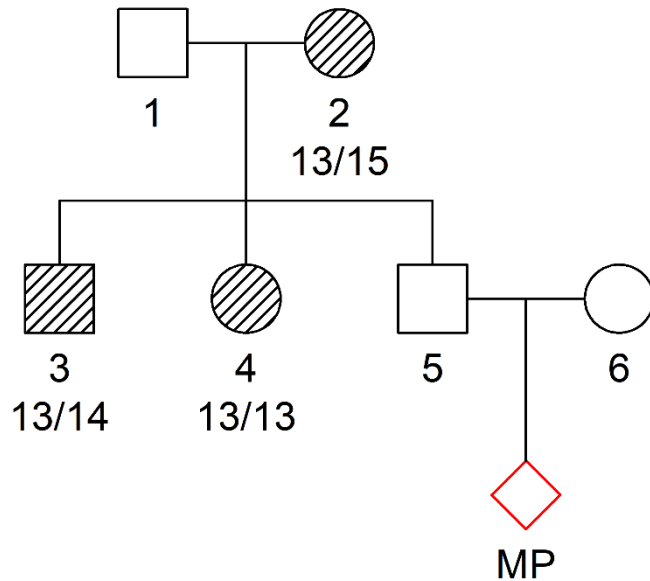
1. Introduction to missing person cases
 - Motivating case: *Missing grandchildren of Argentina*
 - Terminology
 - Genetics
 - Likelihood ratio
 - Software
2. Power
 - Inclusion power
 - Exclusion power
 - Power plots
 - Examples from BNDG

Overview - this lecture

1. Recap from last time
2. Methods for prioritizing in missing person cases
 - Two-step simulation procedure
 - Extended power plots
3. Examples
 - A standard prioritization problem
 - Expanding the marker set
 - X-chromosomal markers
 - Missing great-grandchildren

Missing person cases: Basics

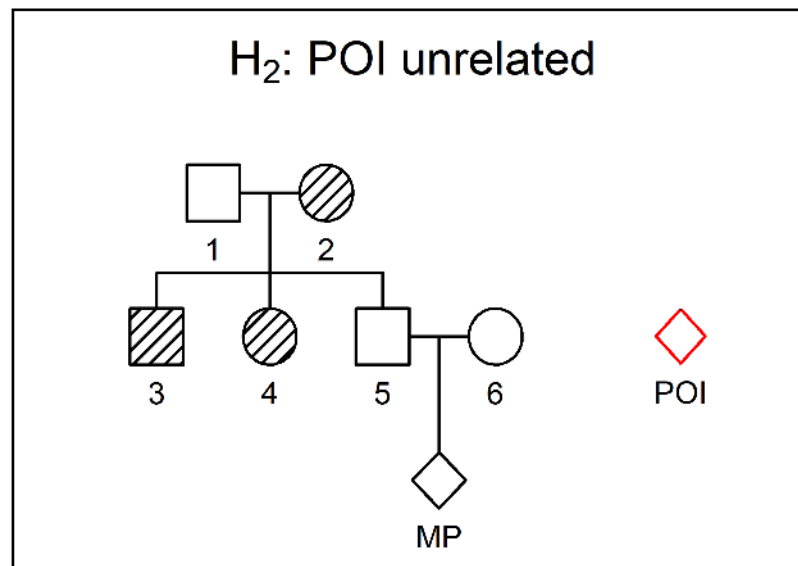
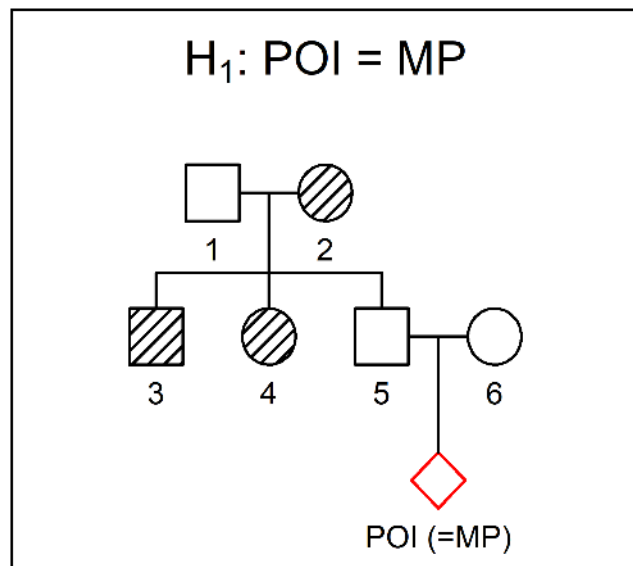
Reference family



Person of interest (POI)



The likelihood ratio (LR)

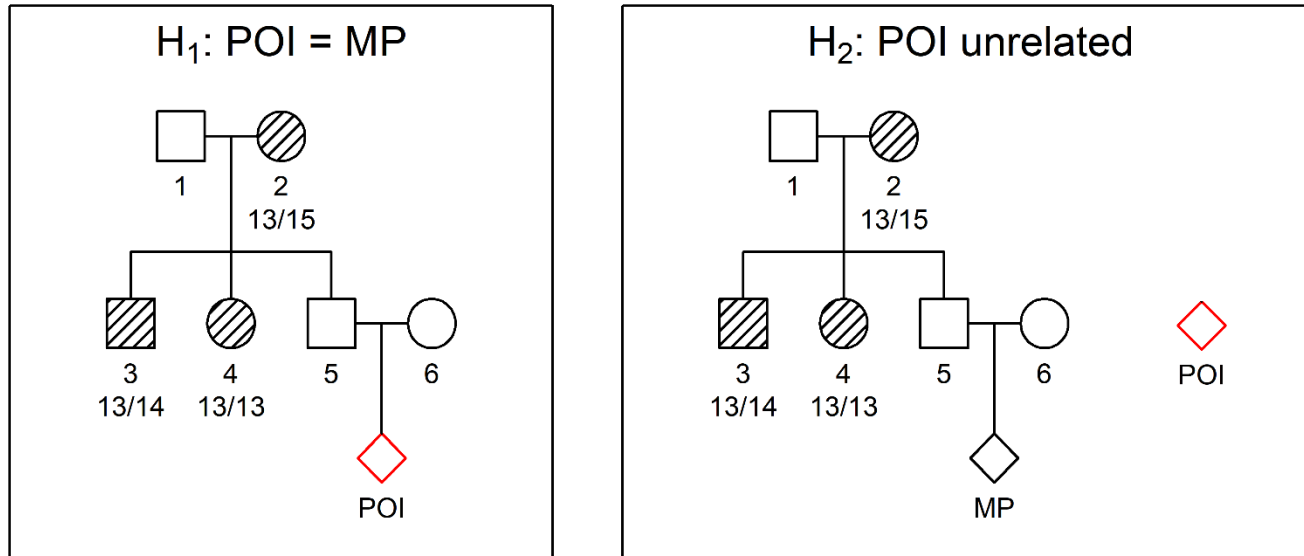


$$LR = \frac{P(\text{data} | H_1)}{P(\text{data} | H_2)}$$

Positive match if $LR > 10,000^*$

*or other suitable threshold

Power in missing person cases

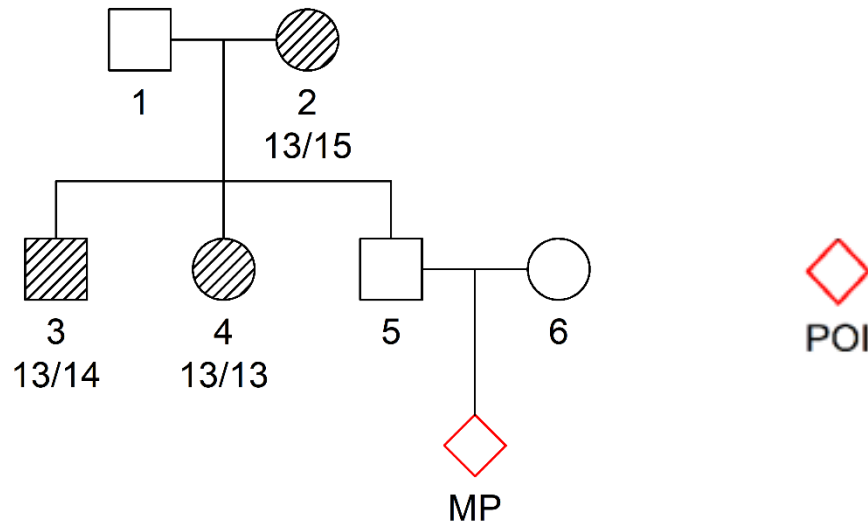


- Two complementary measures of power
 - **Inclusion**: The probability of recognizing the true MP
 - **Exclusion**: The probability of excluding an unrelated POI
- Note: Computed before POI is genotyped!

Depend on

- Reference individuals
- Reference genotypes
- Number of markers
- Allele frequencies

Inclusion power (IP)

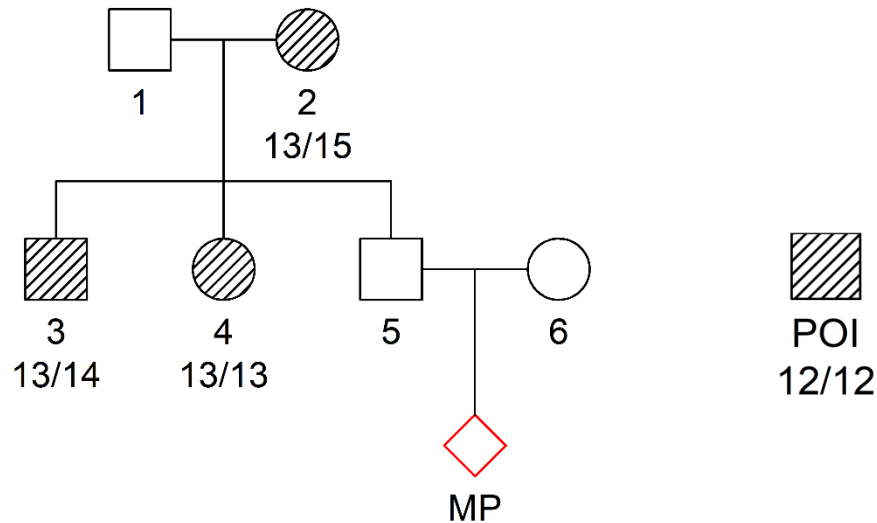


- If **POI = MP**: Do we have enough data to detect it?

$$IP_{10000} = P(LR > 10,000 \mid POI = MP)$$

- Computed by simulation: conditional on reference data

Exclusion power (EP)



POI excluded!
(MP cannot be 12/12)

assumption:
no mutations

- If **POI \neq MP**: What is the probability of mismatch in at least 1 marker?

$$EP = P(\text{exclusion} \mid \text{POI unrelated})$$

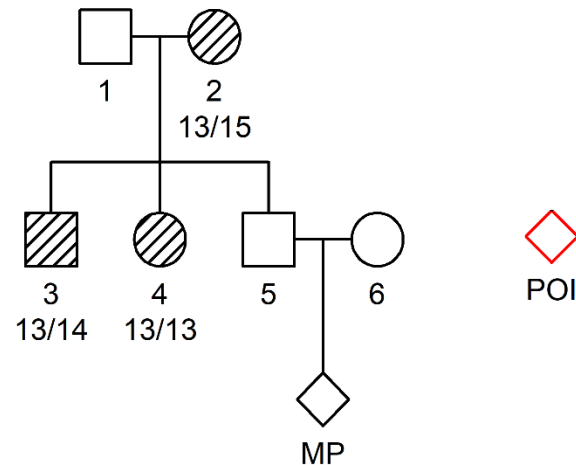
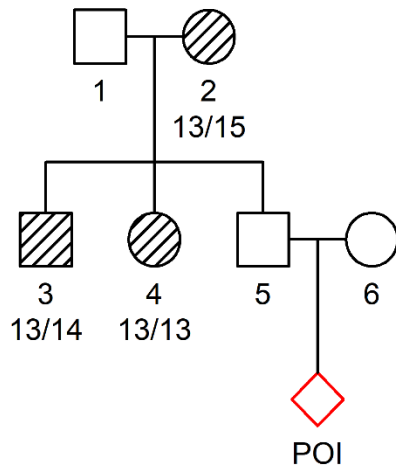
- Can be computed exactly!

The exclusion power formula

NB: Corrected!

Claim: H_1

True: H_2



Single marker: $EP_i = P(\text{mismatch in } H_1 \text{ for marker } i \mid H_2)$

$$= \sum_g I(g \mid H_1) \cdot P(g \mid H_2)$$

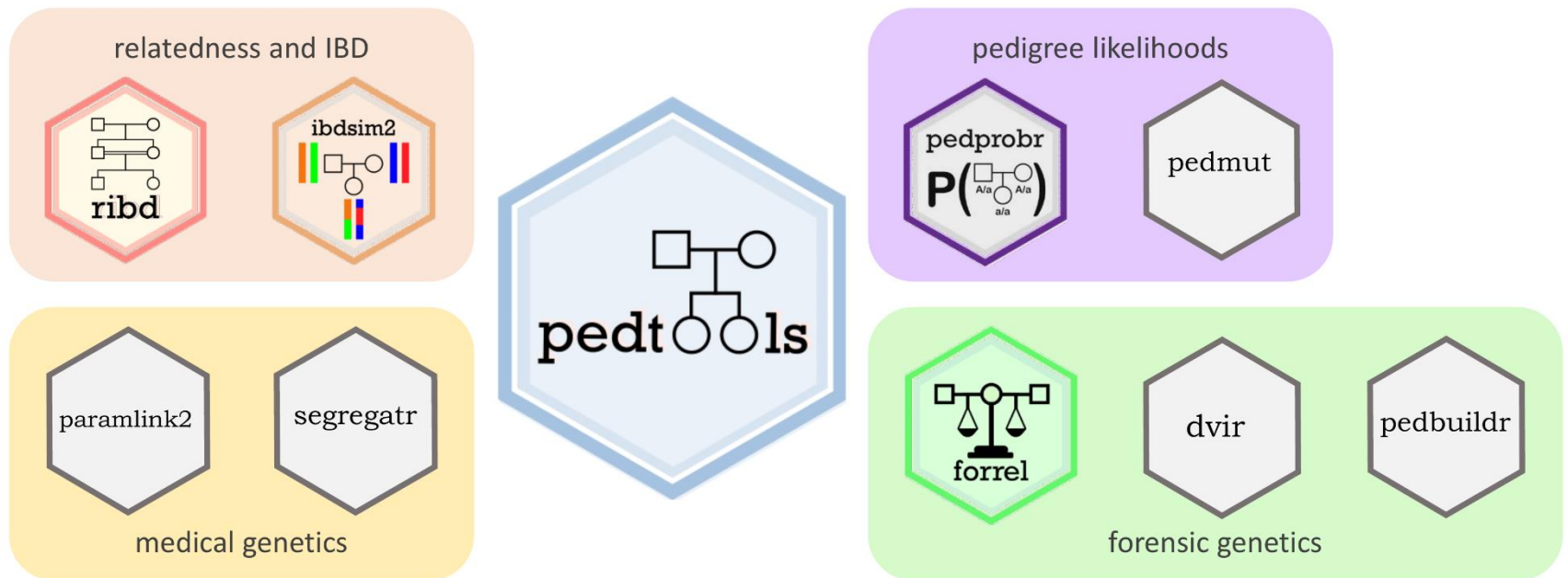
genotype of POI →

$I(g \mid H_1) = \begin{cases} 1, & \text{if } g \text{ incompat with } H_1 \\ 0, & \text{otherwise} \end{cases}$

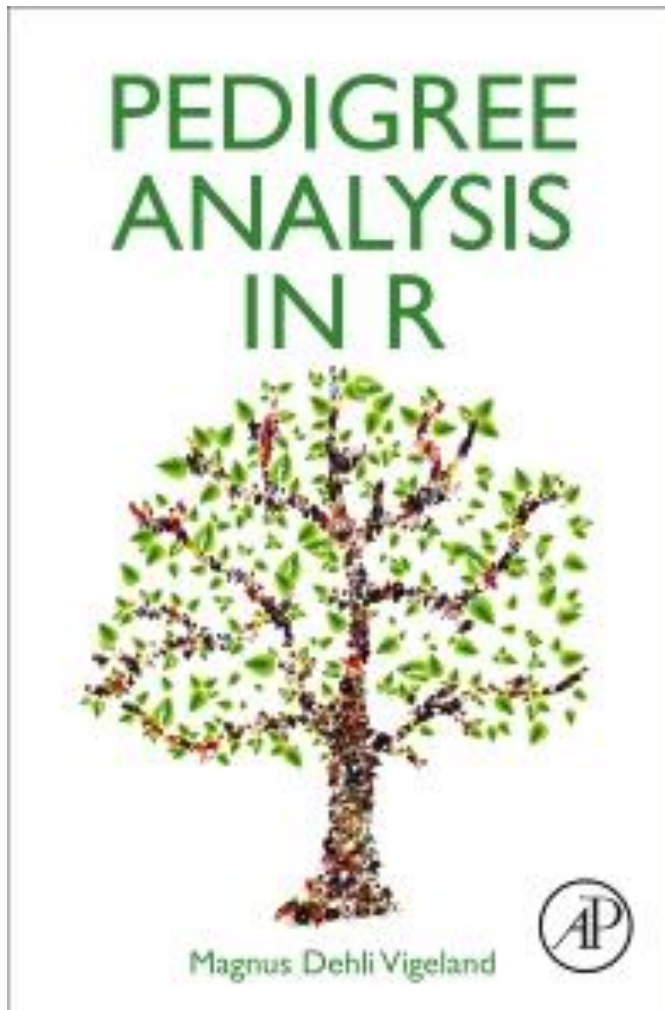
Total power: $EP = 1 - \prod(1 - EP_i)$

The **ped suite**

A collection of packages for pedigree analysis in R



Home page: <https://magnusdv.github.io/pedsuite/>



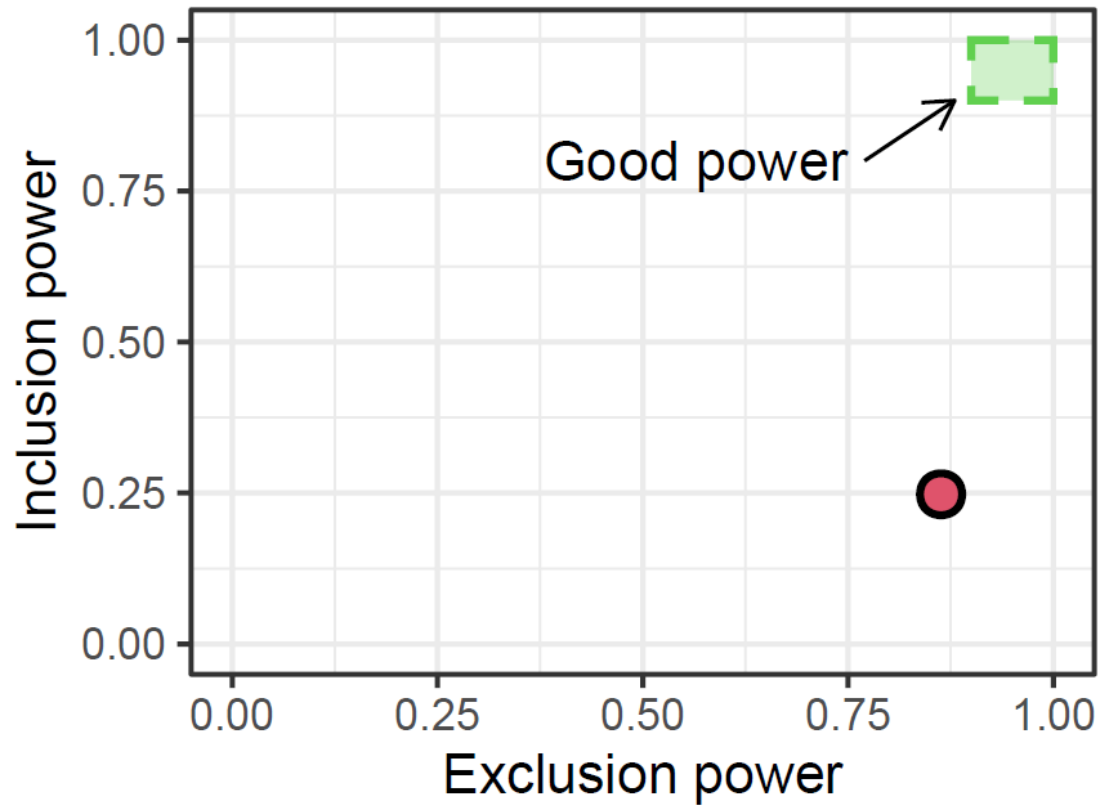
Academic Press, 2021

Available in most online
book stores, Amazon etc.

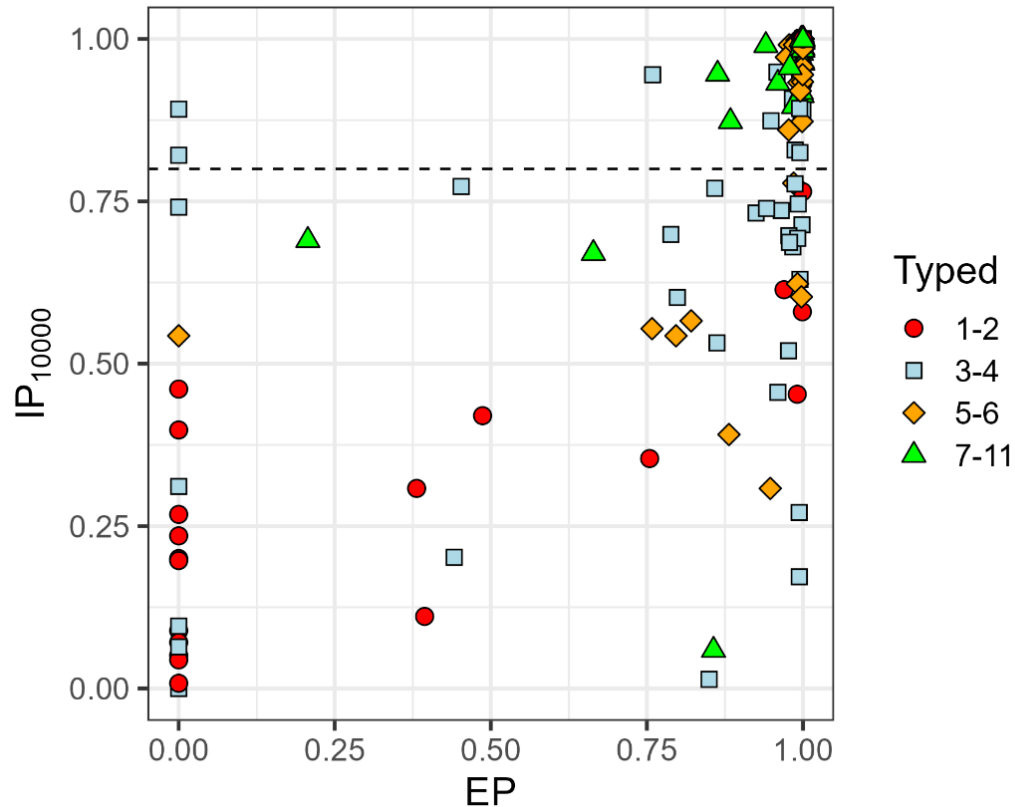
Chapter 6

- Kinship testing
- Missing person cases
- Power calculations
- Visualisations

Power plot



BNDG unsolved cases

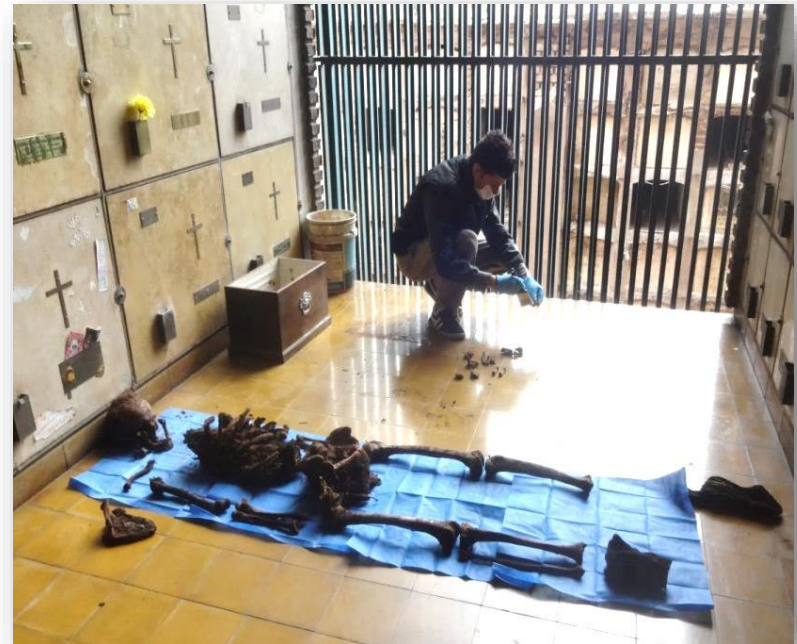
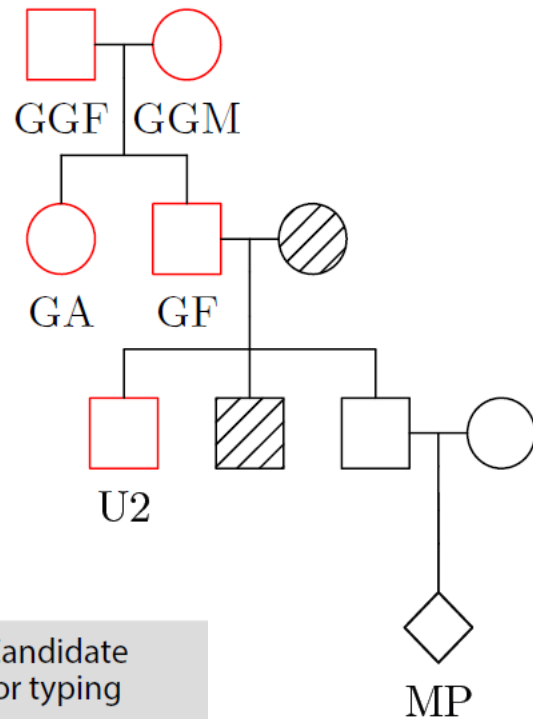


- 34% of cases: **poor power**
- How to improve?
 - retype with more markers
 - recruit additional relatives
 - exhume deceased relatives

Ongoing actions:

- retyping 1000 individuals
- large-scale exhumations

The prioritization problem





Contents lists available at [ScienceDirect](#)

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen



Research paper

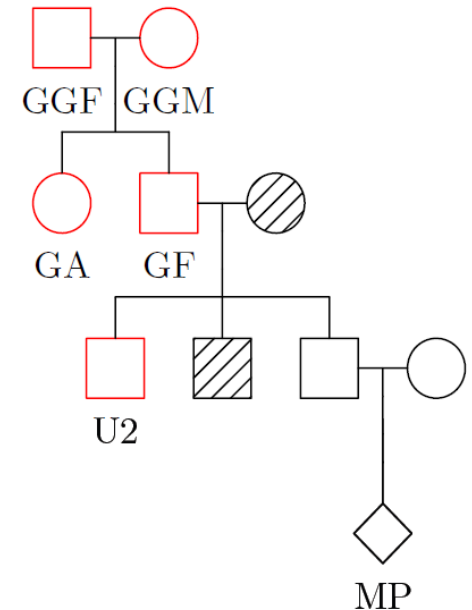
Prioritising family members for genotyping in missing person cases: A general approach combining the statistical power of exclusion and inclusion

Magnus D. Vigeland^{a,*}, Franco L. Marsico^b, Mariana Herrera Piñero^b, Thore Egeland^c

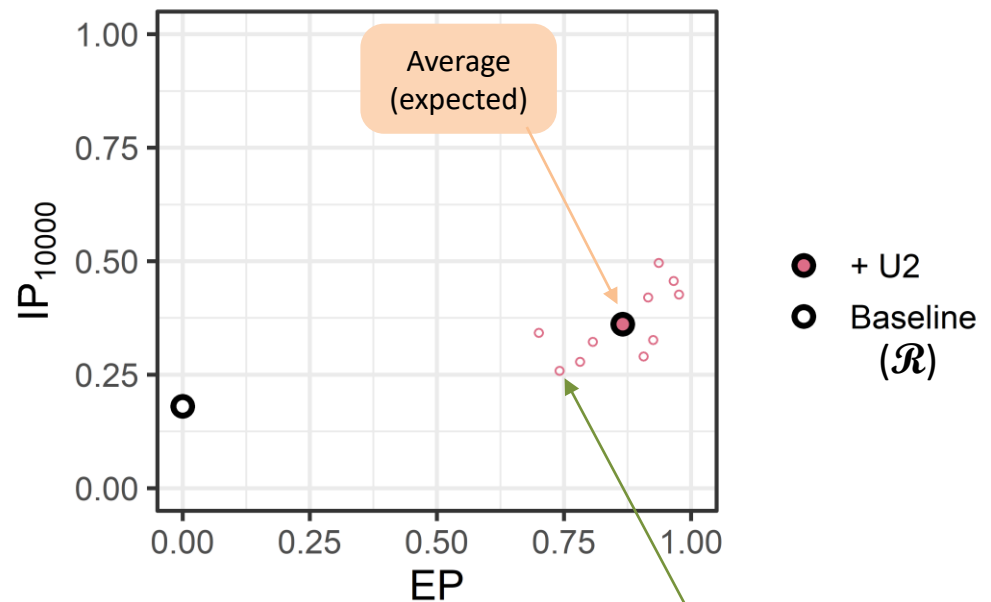
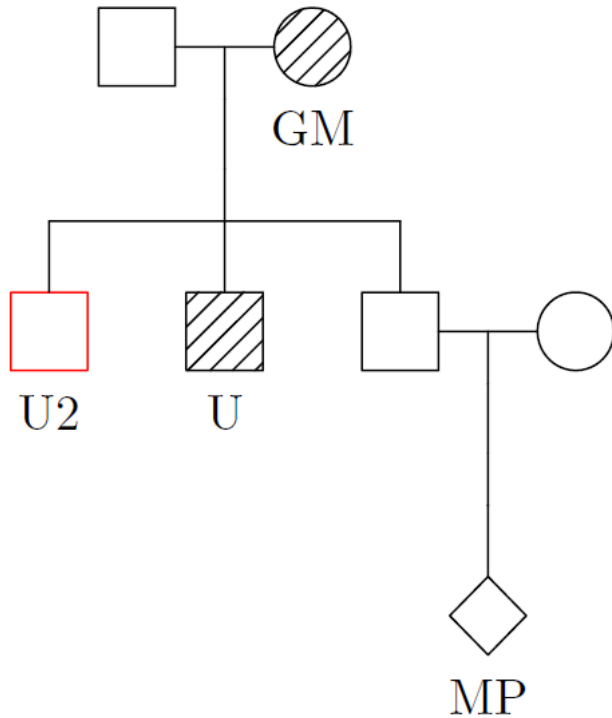


Simulation procedure

- Input:
 - Reference data \mathcal{R}
 - List of possible additions S_1, S_2, \dots
- For each S_i :
 - simulate p profiles for S_i conditional on \mathcal{R}
 - for each sim:
 - estimate IP by simulation of MP (e.g. $q = 1000$ sims)
 - compute EP
- Display results in power plots

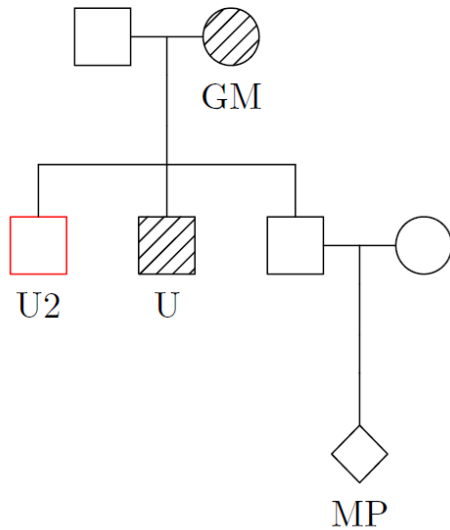


Example



- Baseline reference $\mathcal{R} = U + GM$ (23 markers)
- 10 simulations of U2 conditional on \mathcal{R}

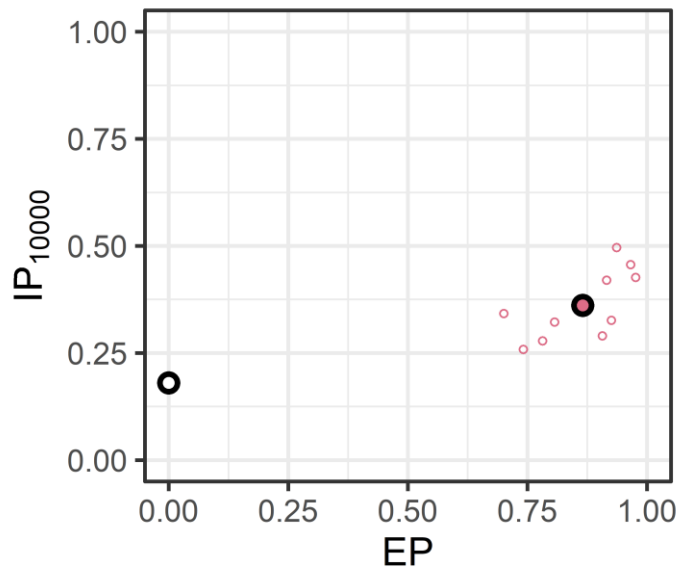
Example (cont.)



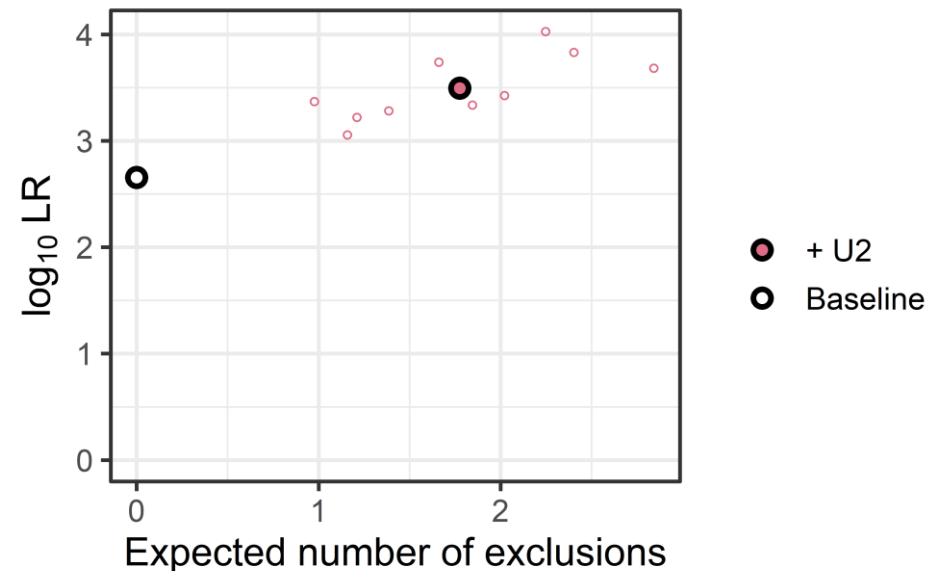
Complementary power plot

- Expected number of exclusions
(i.e., #markers with mismatch)
- Expected $\log(\text{LR})$

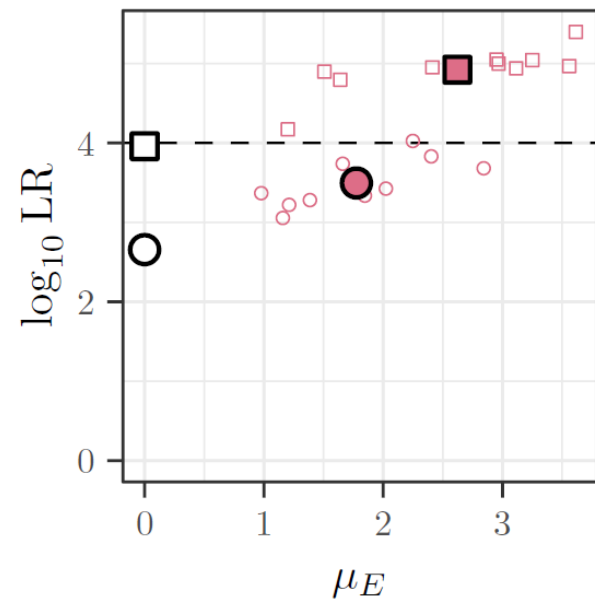
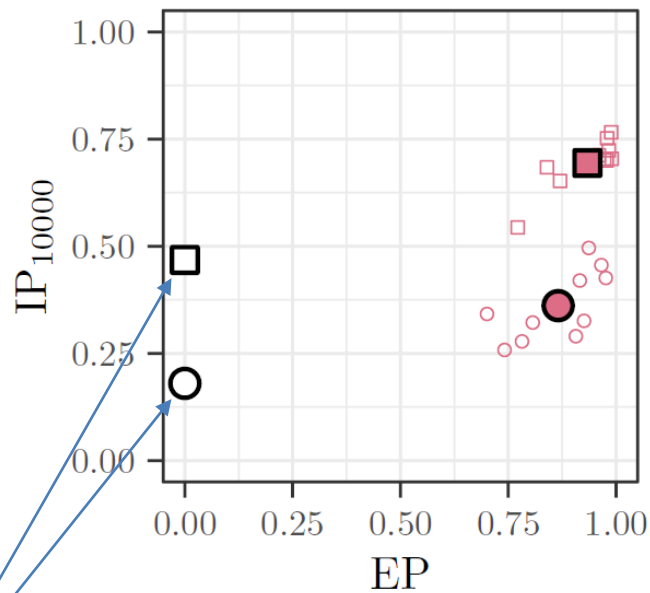
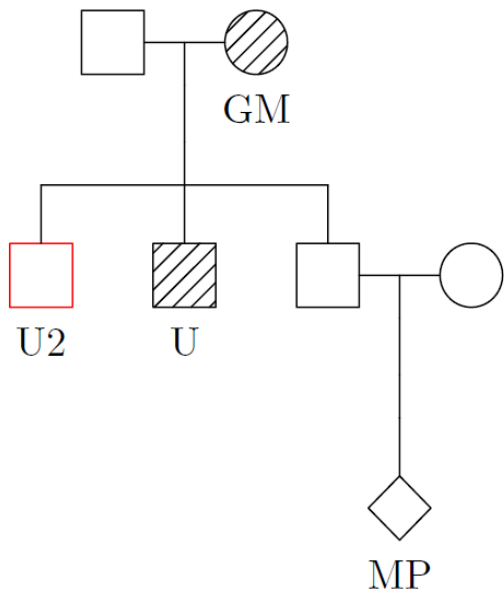
Power plot 1



Power plot 2

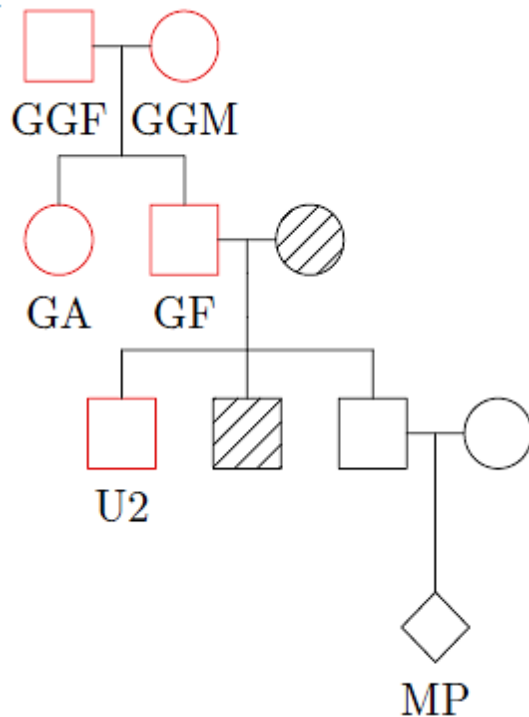


The importance of conditional sims



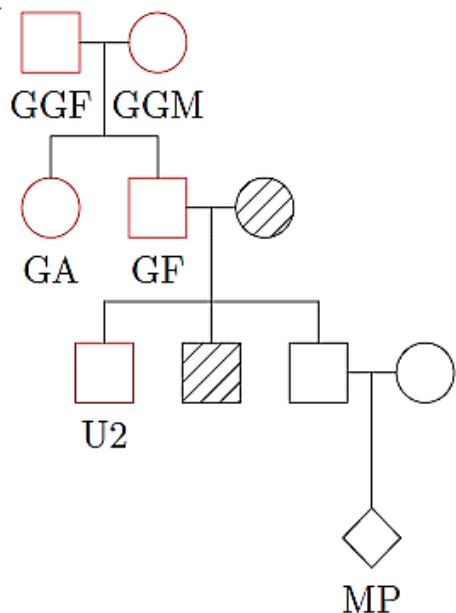
Different
baselines

A larger example



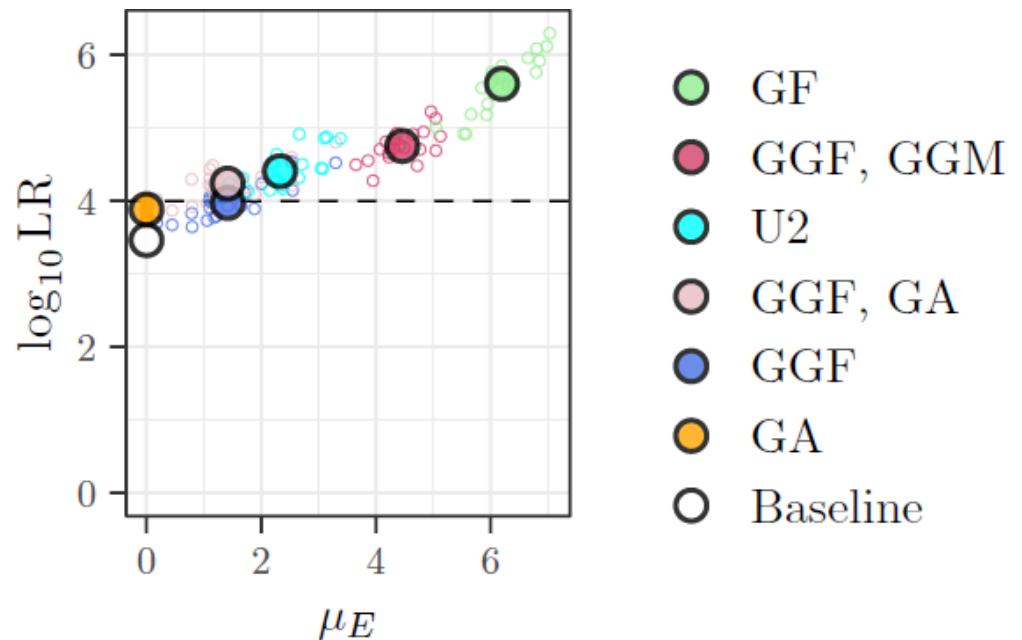
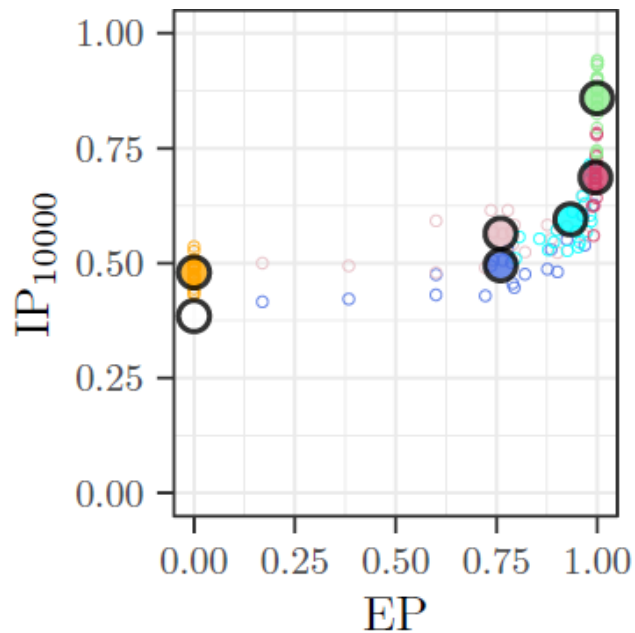
Questions

- Who is most informative?
- Are GGF + GGM as informative as GF?
- Who is most informative of GA and GGF?
- What about U2?

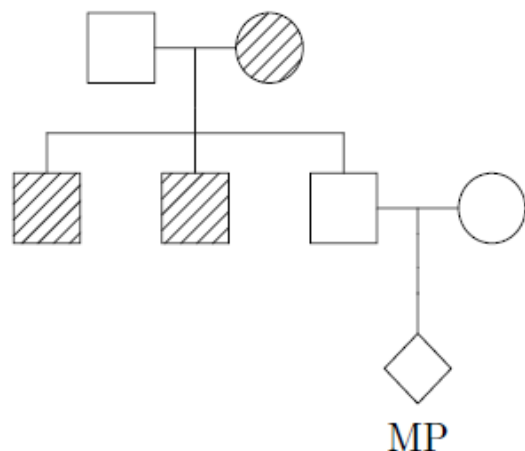


Questions & answers

- Who is most informative?
→ GF
- Are GGF + GGM as informative as GF?
→ Not quite
- Who is most informative of GA and GGF?
→ GGF (gives exclusion!)
- What about U2?
→ Slightly worse than GGF+GGM



Expanding the marker set



Originally typed with 15 markers.

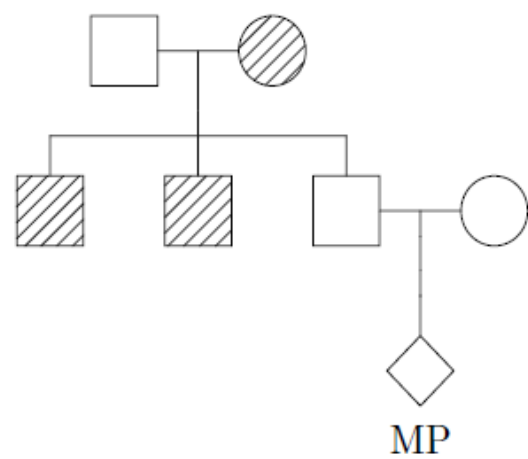
$$IP \approx 0.05$$

$$EP \approx 0.50$$

How much would it help to add more markers?

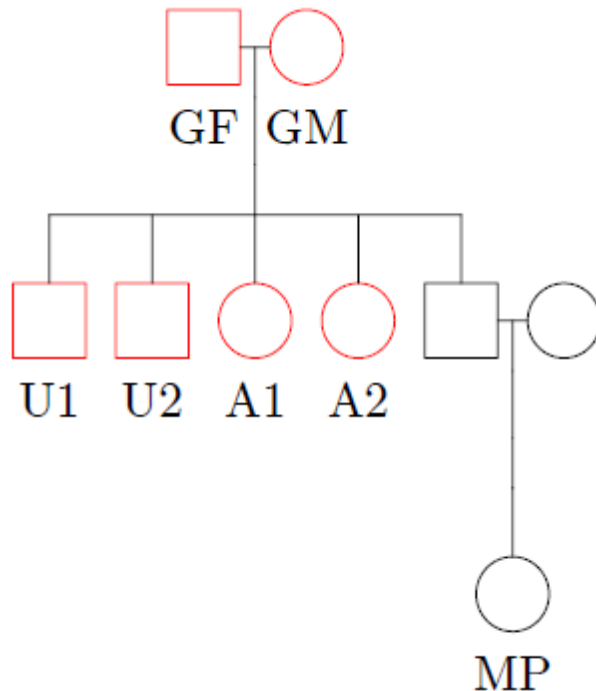
- Set1 (15 markers): CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, TH01, TPOX, VWA.
- Set2 (23 markers): Set1 + D1S1656, D2S441, D10S1248, D12S391, D22S1045, PENTA_D, PENTA_E, SE33.
- Set3 (33 markers): Set2 + D2S1360, D3S1744, D4S2366, D5S2500, D6S474, D6S1043, D7S1517, D8S1132, D10S2325, D21S2055.

A



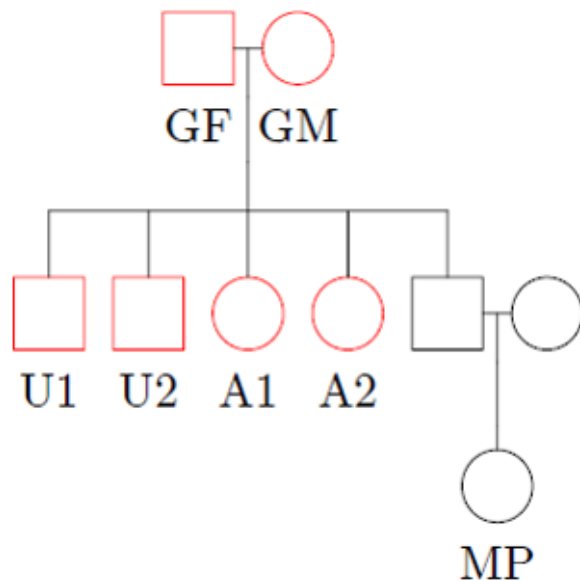
X-chromosomal markers

Reference

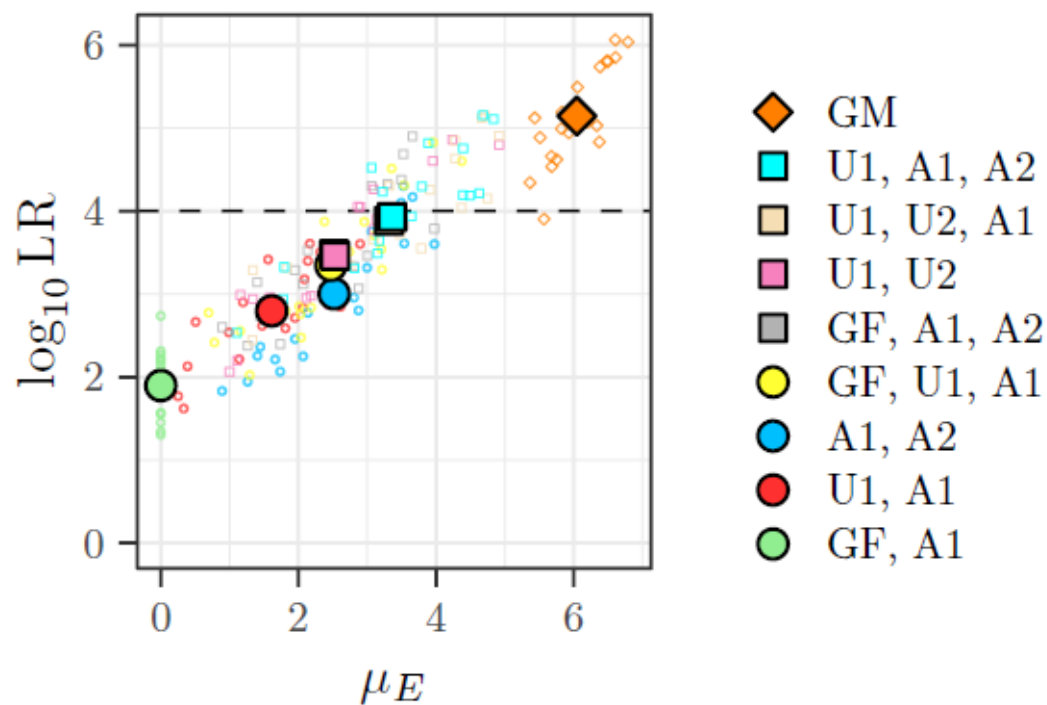
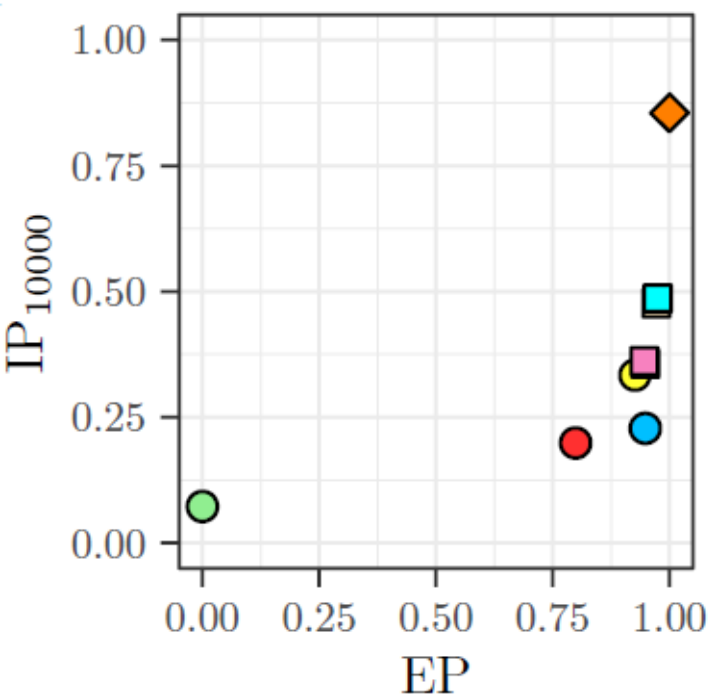


- Are X-chromosomal markers useful here?
- Who are most informative?
- Who are least informative?

- Argus X-12 kit (Qiagen)
- 12 STR markers on X
- Linkage and LD: Yes, but will not affect main conclusions

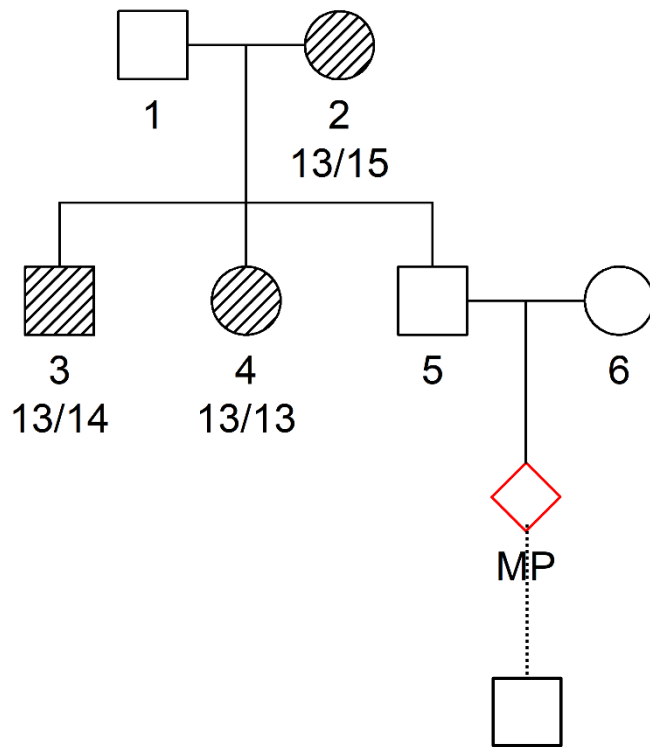


Grandma



Missing **great**-grandchildren

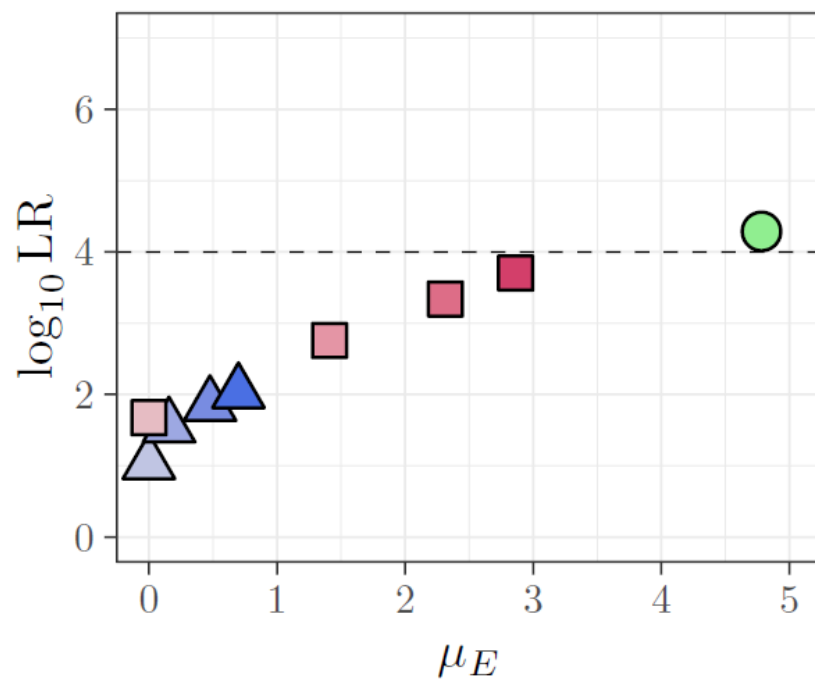
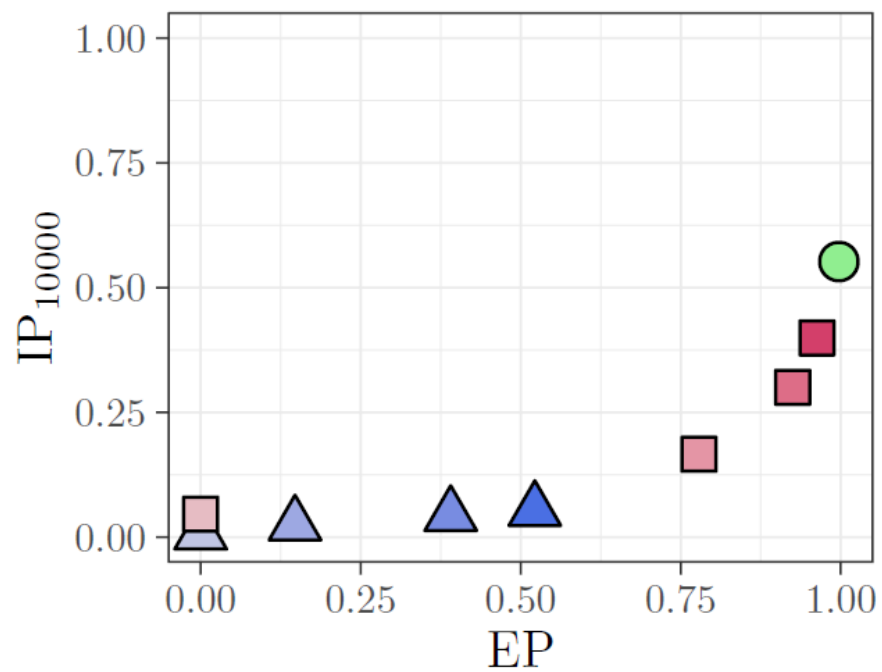
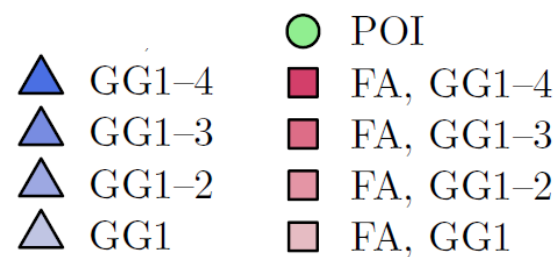
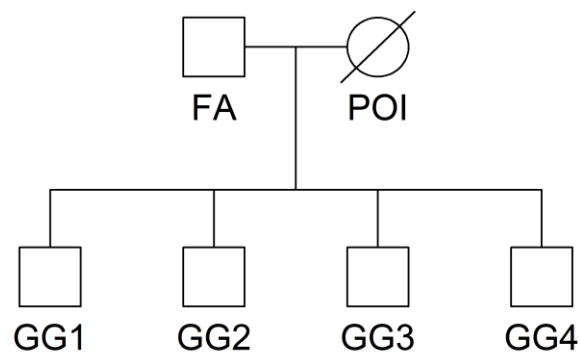
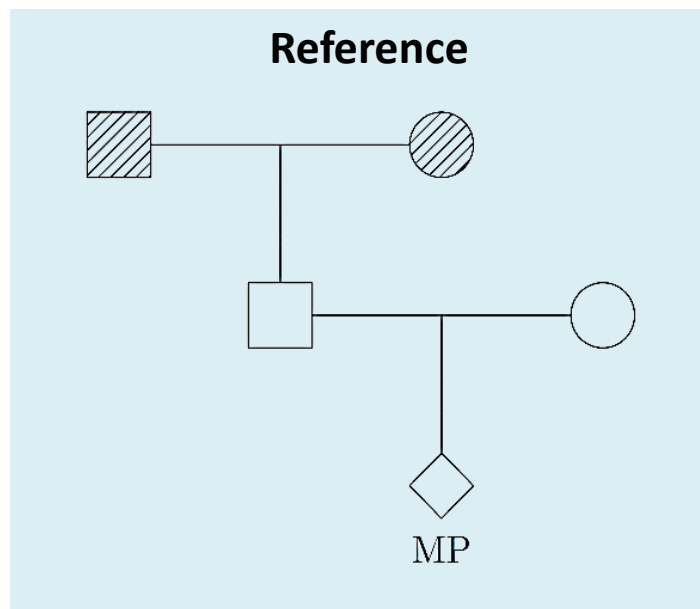
Reference family

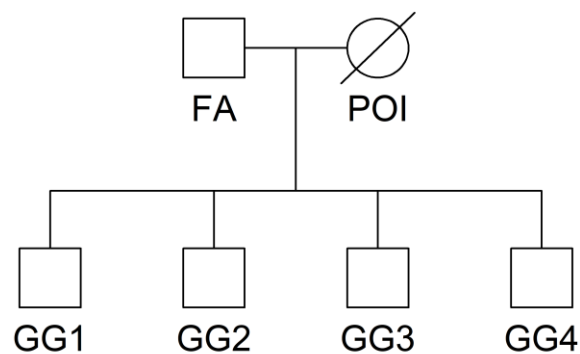
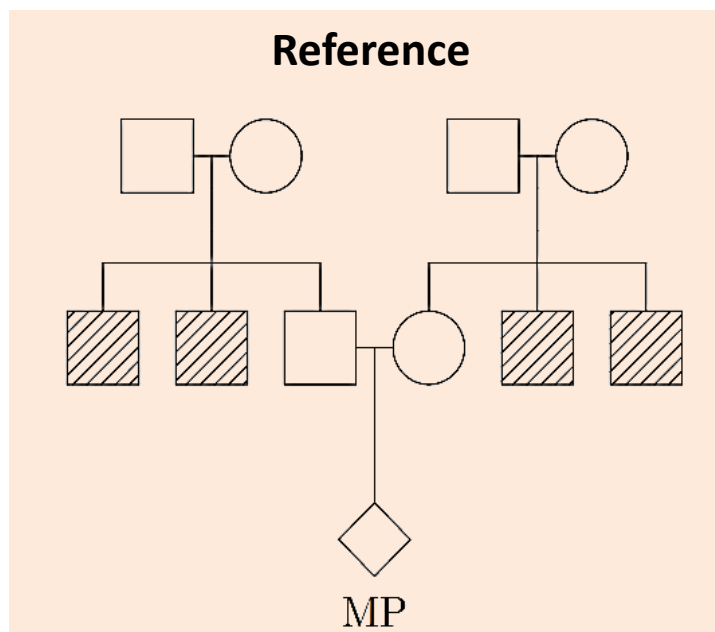


match?

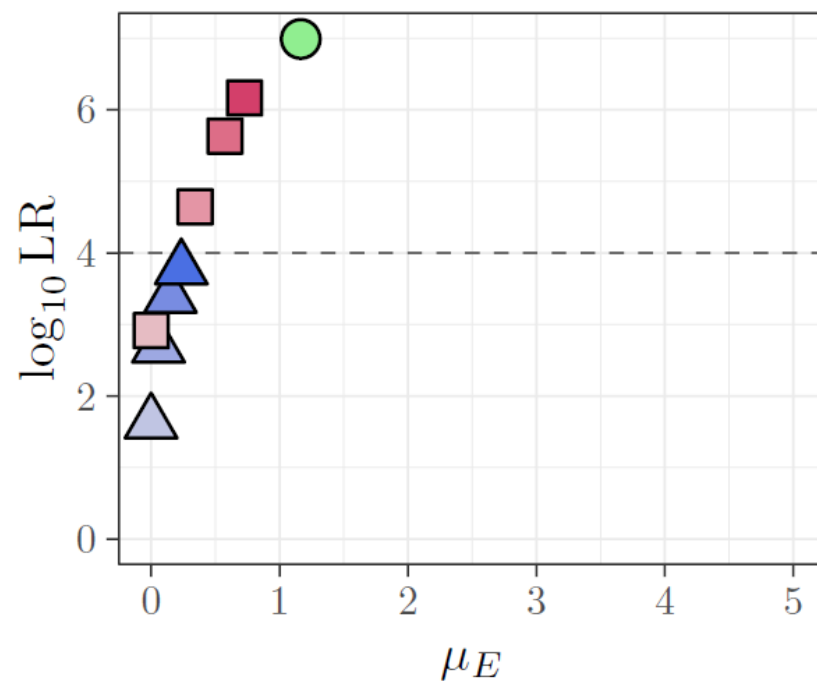
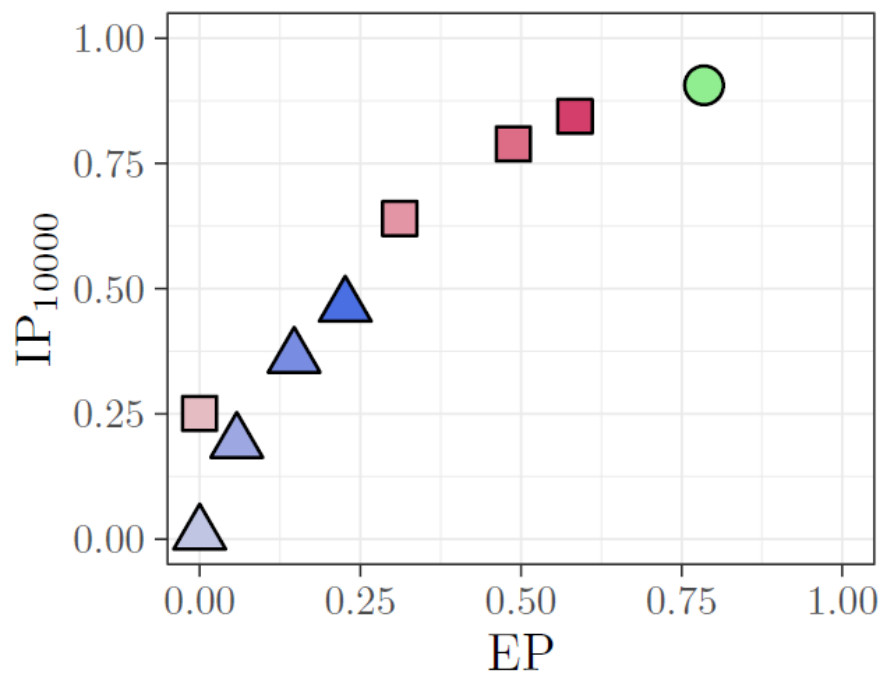
I think my mother was MP!





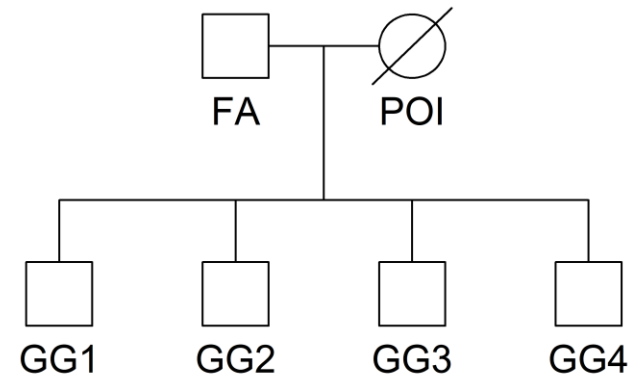


- | | | | |
|--|-------|--|-----------|
| | POI | | FA, GG1-4 |
| | GG1-4 | | FA, GG1-3 |
| | GG1-3 | | FA, GG1-2 |
| | GG1-2 | | FA, GG1 |
| | GG1 | | |



Great-grandchildren: Lessons from cases

- Single great-grandchild: **generally poor power**
 - exception: If parent(s) of MP are typed
- **The other parent** boosts the power!
- Optimally, include
 - at least 2 great-grandchildren
 - AND their (other) parent
- This is almost as good as POI herself



Thank you!