DNA Mixture Interpretation Principles: Observations from a NIST Scientific Foundation Review
AAFS 2019 Workshop #10 (February 18, 2019; Baltimore, MD)

Approaches to Defining Limitations

Keith Inman
Leverhulme Centre for Forensic Science
University of Dundee
Criminal Justice Department
Cal State East Bay

One thing is not the other

- Sample/Item
- Sampling
- Extract
- Quant
- Amplification
- Detection
- Display

Steps in DNA Analysis and Interpretation

swab

Gathering the Data

Understanding Results Obtained

Collection/Storage/ Characterization → Extraction/ Quantitation → Amplification/ Marker Sets → Separation/ Detection → Data → Stats → Report → result

Interpretation

Probabilistic Genotyping Software (PGS)

What are the relevant questions?

- Common wisdom (mythology):
  - Using more information in model present in the typing improves performance (>LR)
  - Question seems to be: how much can we model?
- Perhaps a better question:
  - What is the source of variation within any amplified DNA?
  - Remembering that the sample is not the evidence

Focus has been on software, not sample itself…

Creating a mixture set: Complexity variables for *Sample*

- Created a set of complex mixtures
  - 2, 3, and 4 person mixtures
- Mixture ratios
  - 1:1, 2:1, 4:1, 9:1
- Total DNA
  - 500, 100, 50, 30 pg total
  - Highest amount of DNA in any mixture: 250 pg
  - Lowest amount of DNA in any mixture: 3pg (half of a diploid cell)

Number of samples created

- Combinations of conditions = 164
- Replicates = 5
- Total samples = 820

Methodology

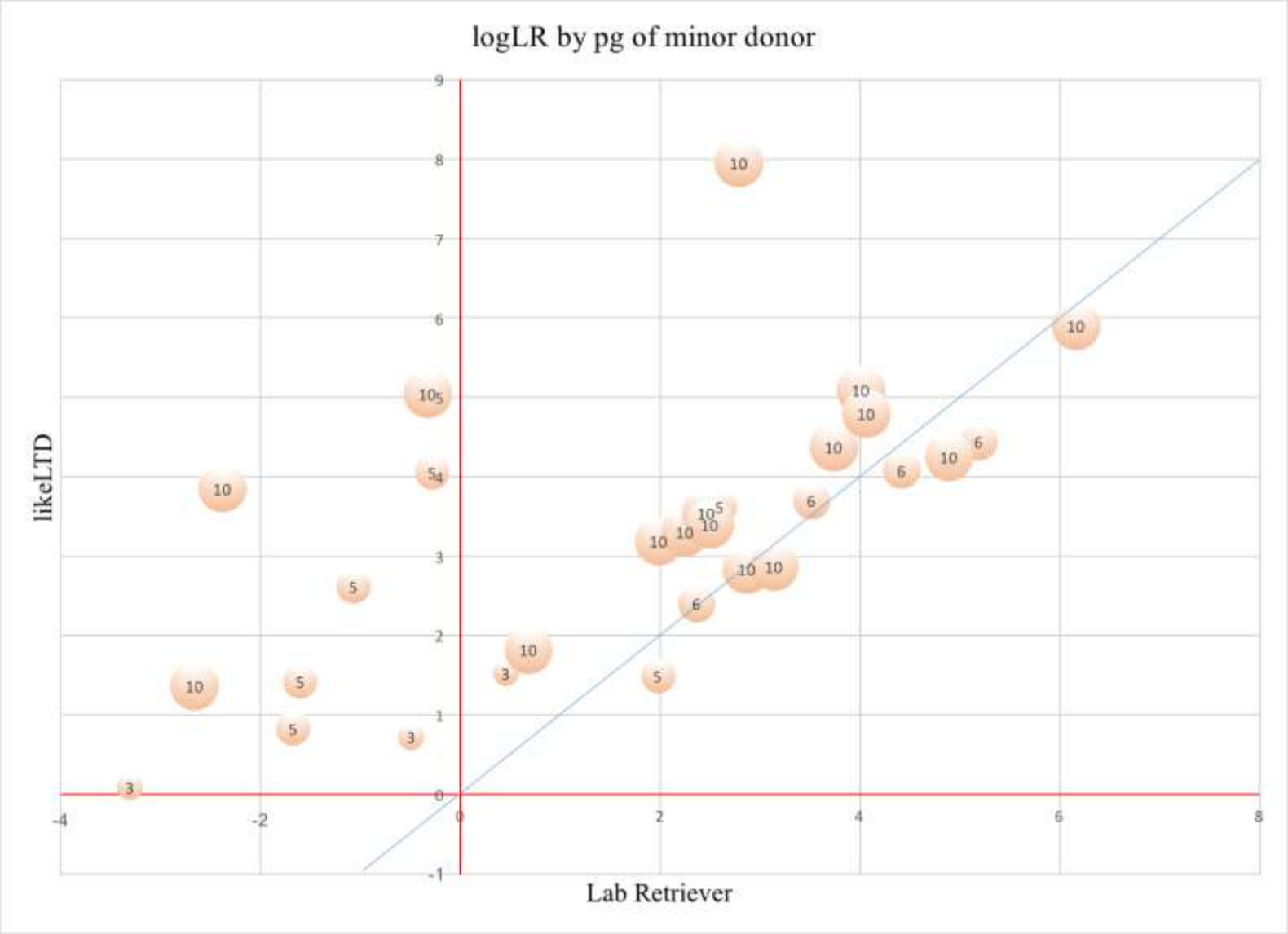
- Analyze all samples with an analytical threshold determined on a per-run basis
  - This varied by color and run
  - Min 10 RFU
  - Max 30 RFU
- Run LR with
  - *Discrete* variable systems (**Lab Retriever** and **LRMix**) on ***minor*** donor
  - *Continuously variable* systems (**likeLTD** and **European Forensic Mixtures**)
  - All open source

Evidential Efficiency

- Evidence against a suspected contributor can never be stronger than the ***inverse match probability*** for that contributor obtained by a single source DNA profile
- A mixed sample can never give stronger evidence than a high-quality single source profile
- Standardized the LR's against the RMP
  - Comparing the efficiency of the programs relative to the maximum amount of information that can be derived from the reference

    - Cowell, et.al; 2013

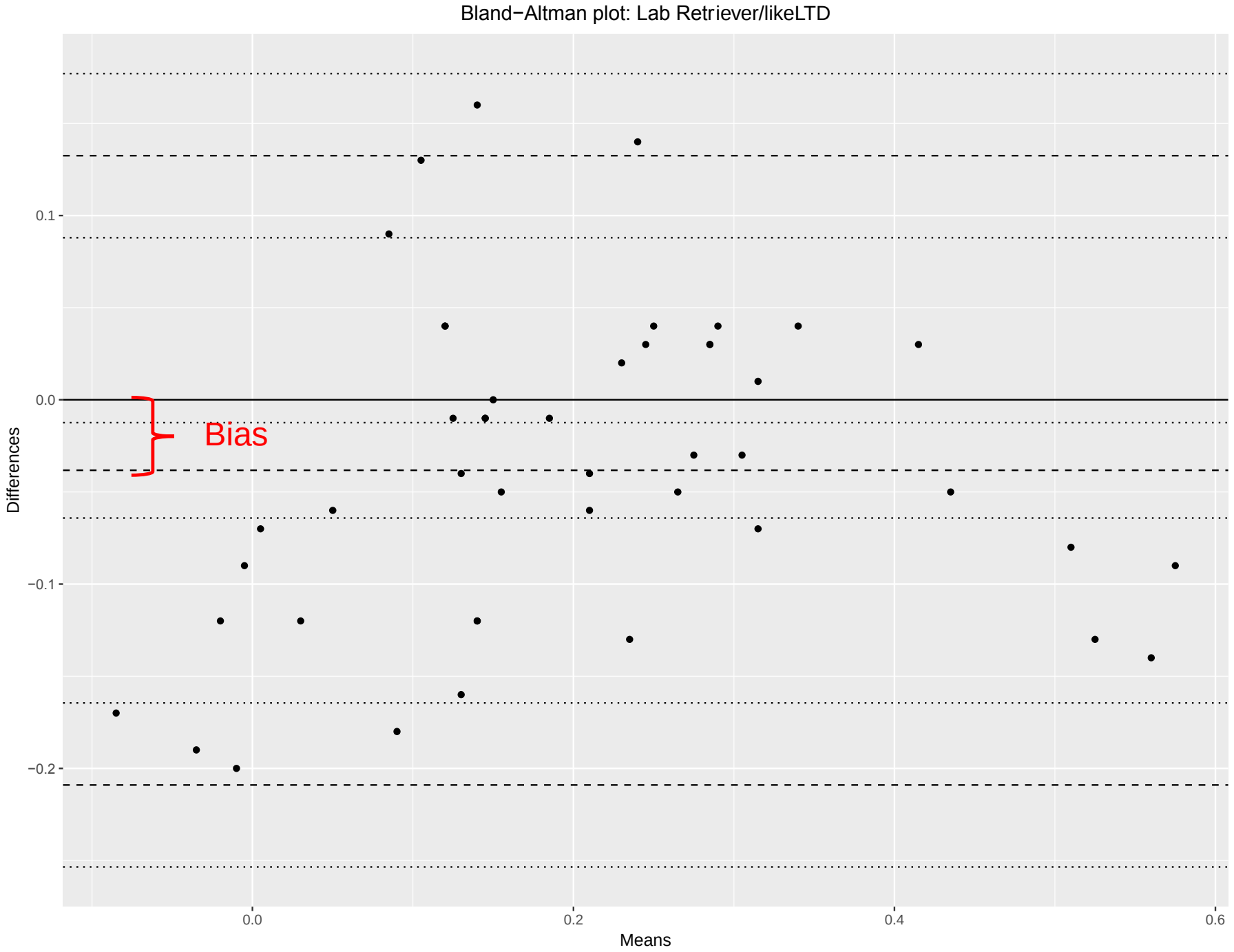
Testing first the difference in models



Model 1: Peaks only

Model 2: Peak Hts

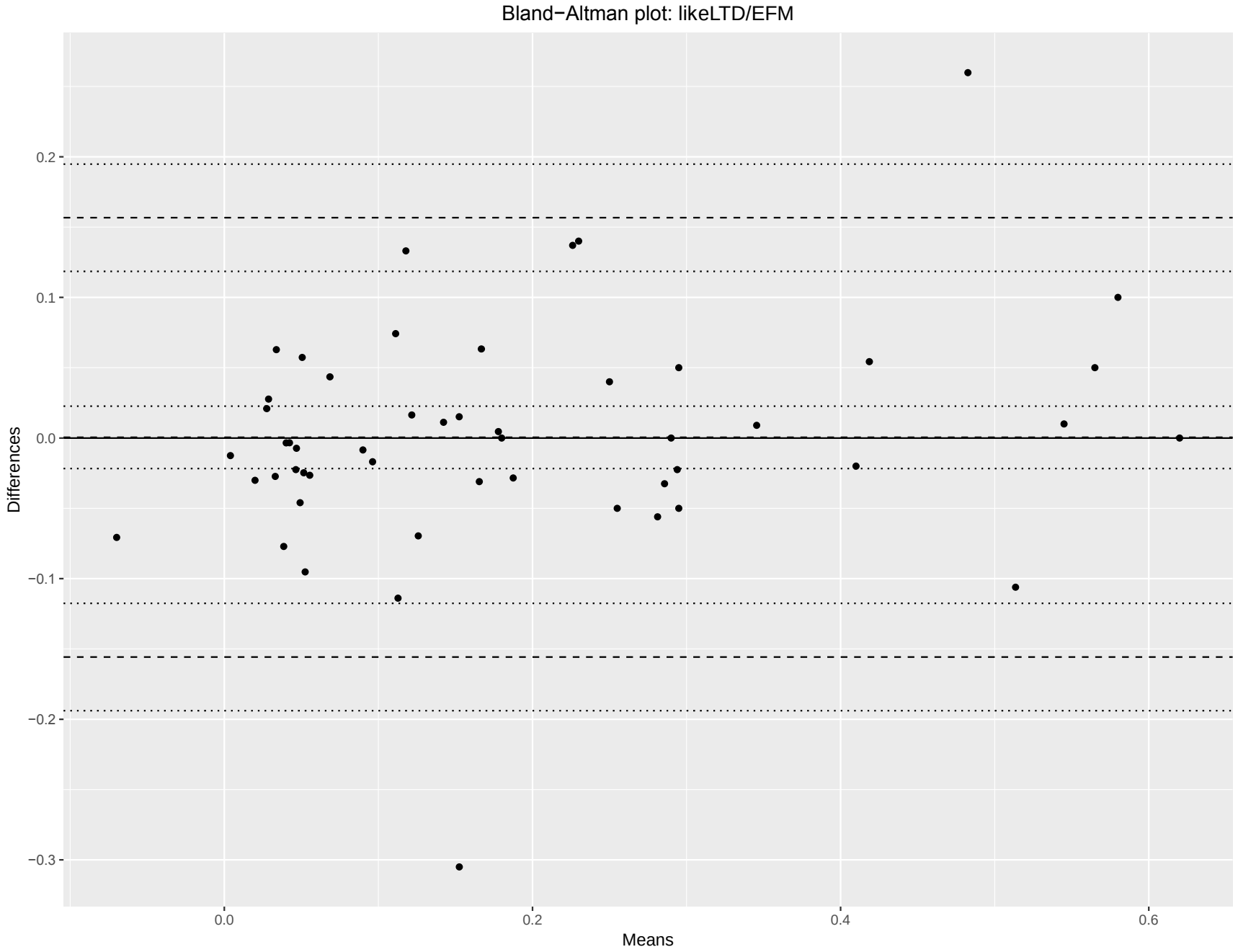
Bias present



Model 1: Peaks Hts

Model 2: Peak Hts

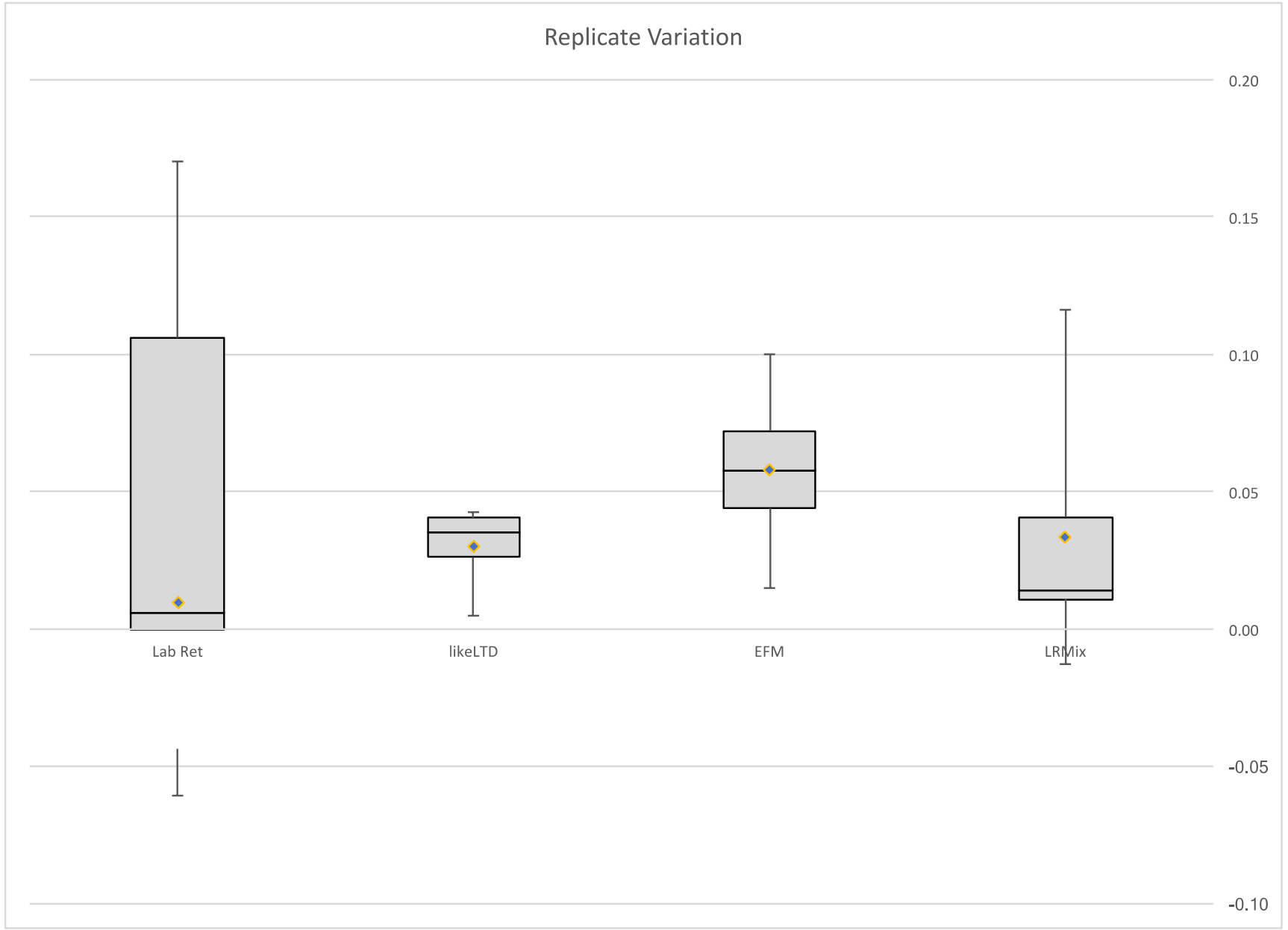
No bias evident



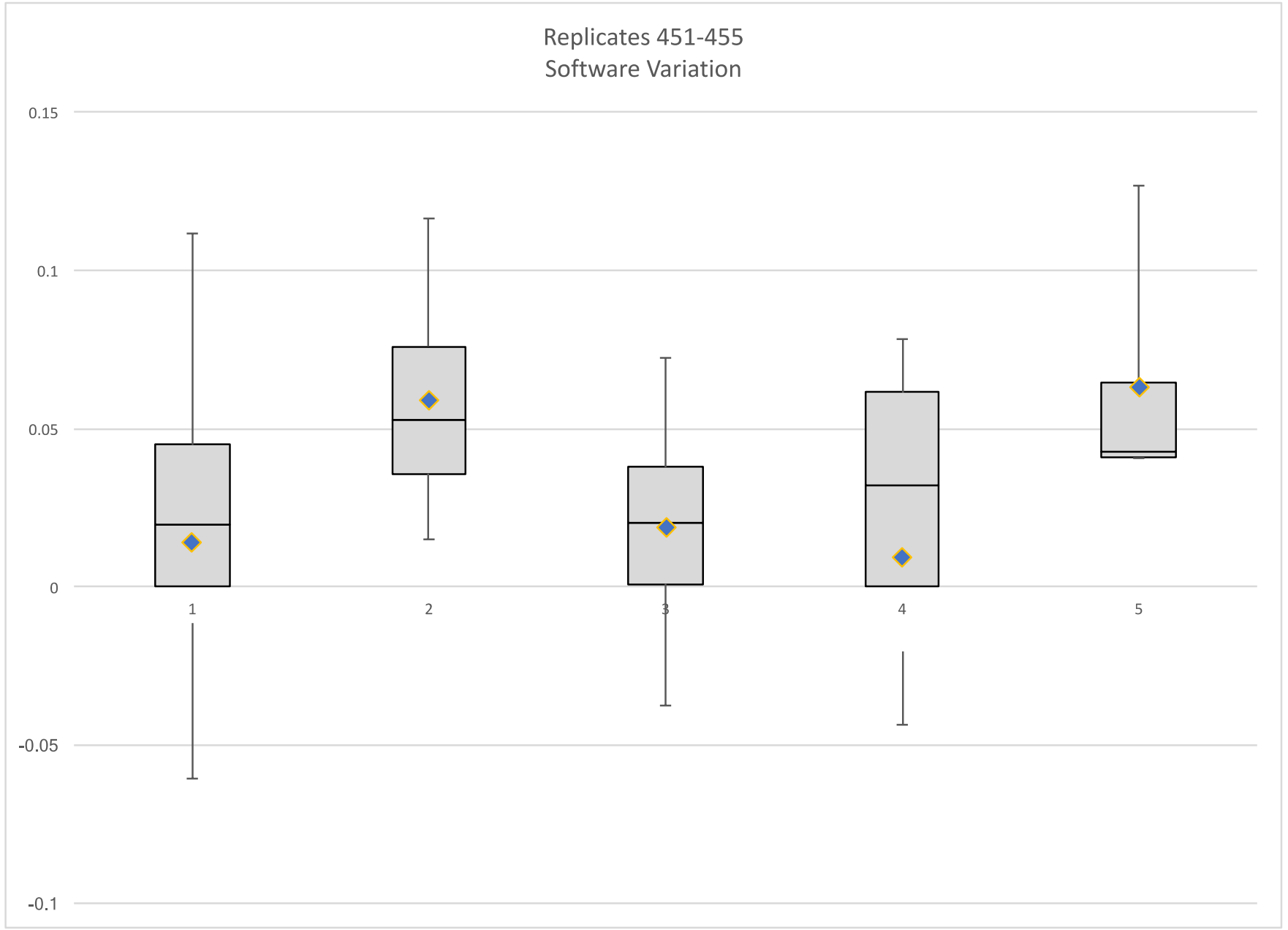
So far, conventional wisdom upheld

- Incorporating more analytical information provides ‘better’ LR
- But what about complexity inherent in the sample?
  - Use the software to explore sample complexity
    - The electropherogram is NOT the bloody trousers

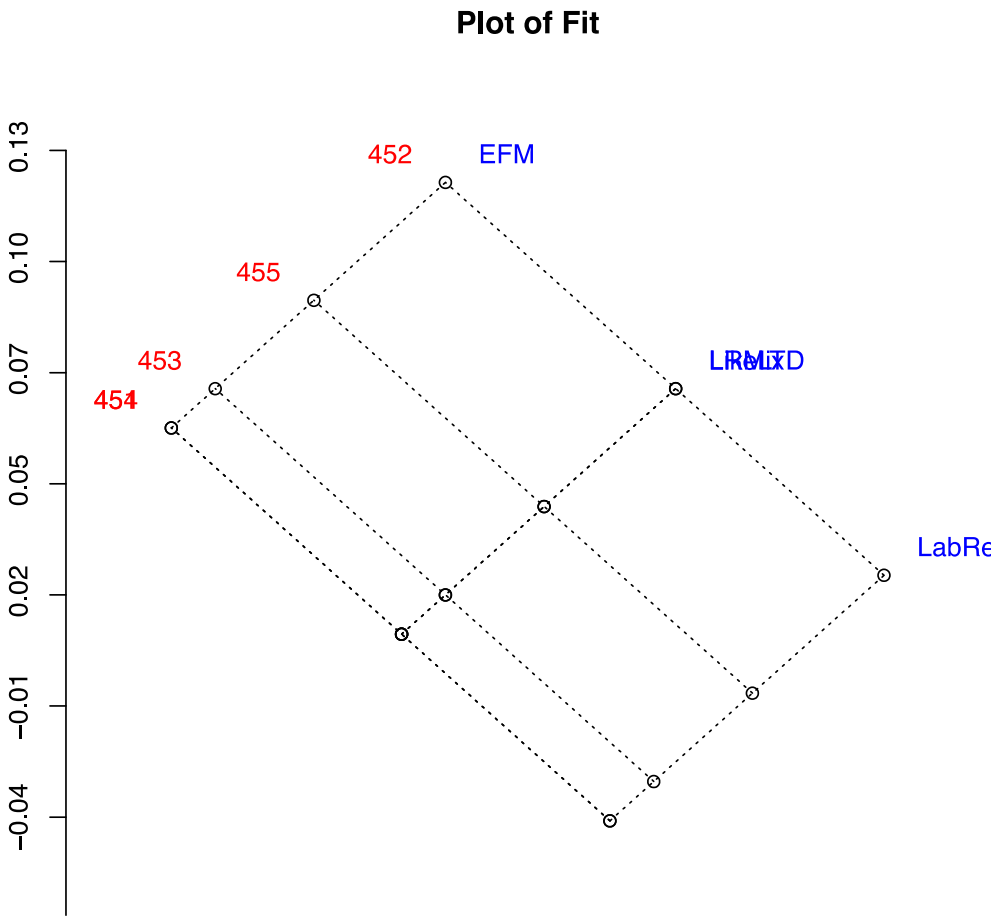
50pg 1:1


Replicate Variation
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
Lab Ret
likeLTD
EFM
LRMix

50pg 1:1

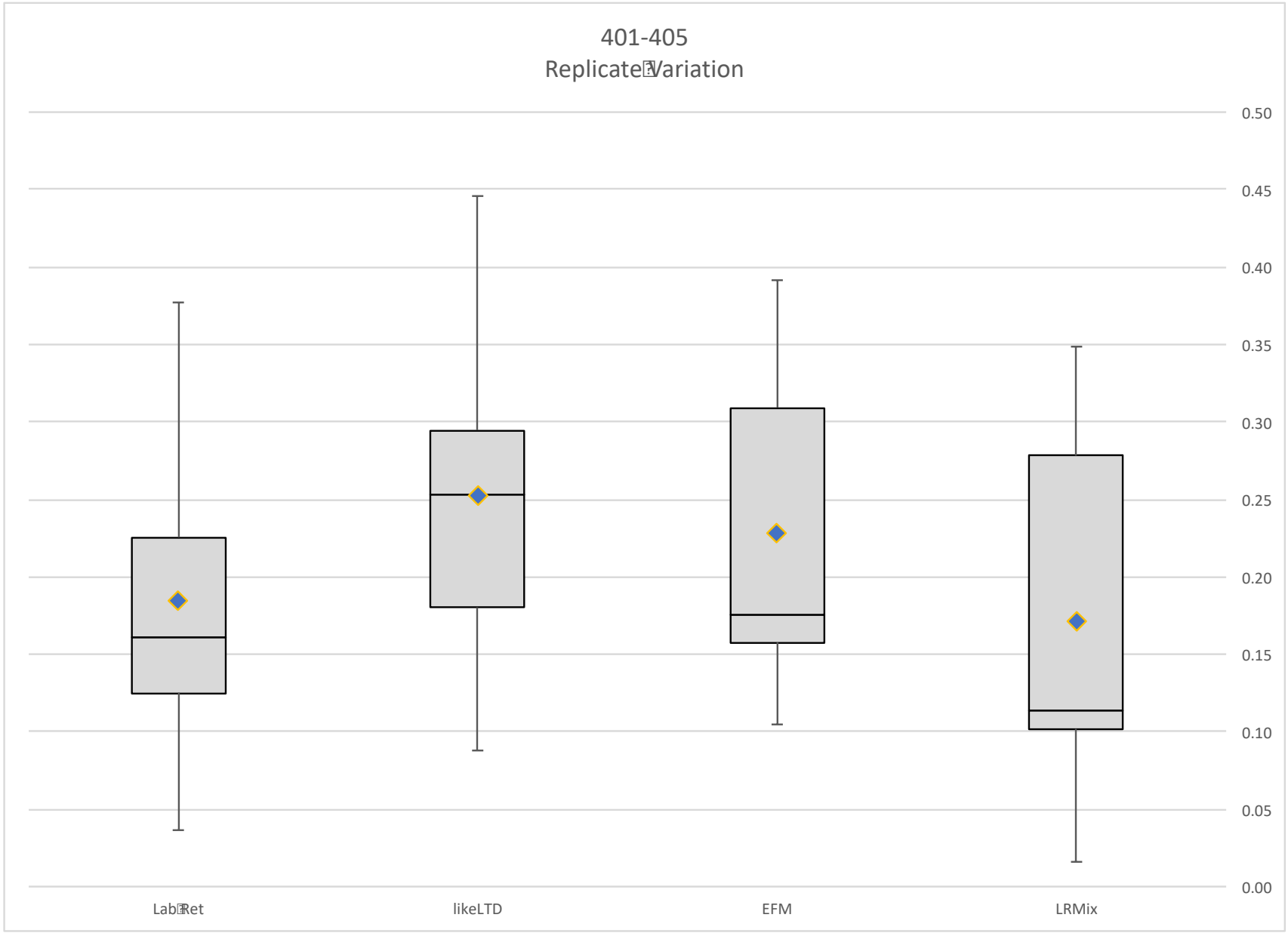

Replicates 451-455
Software Variation
0.15
0.1
0.05
0
-0.05
-0.1
1
2
3
4
5

Median polish of replicates

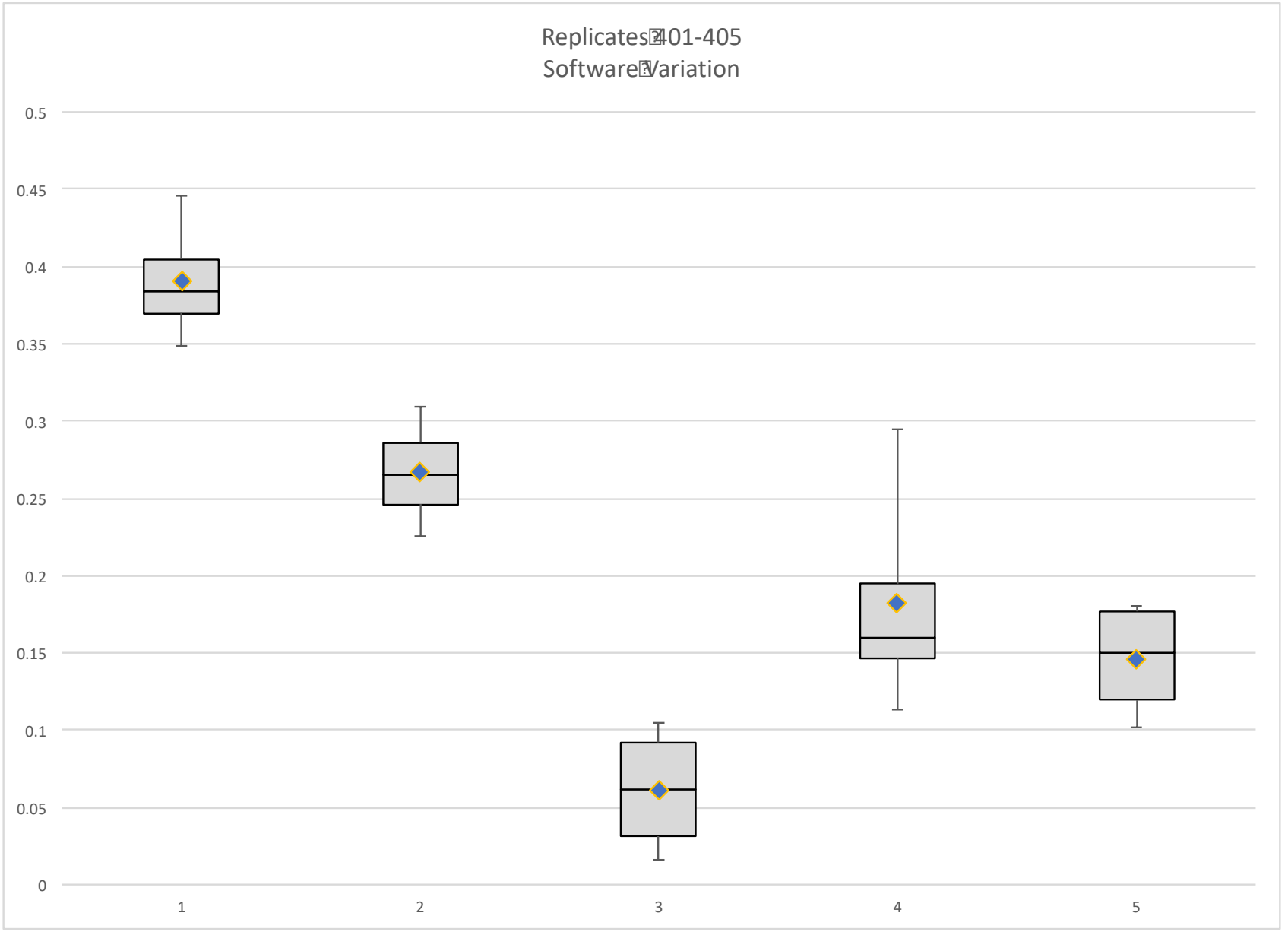


SDreps/SDalgs 0.6543

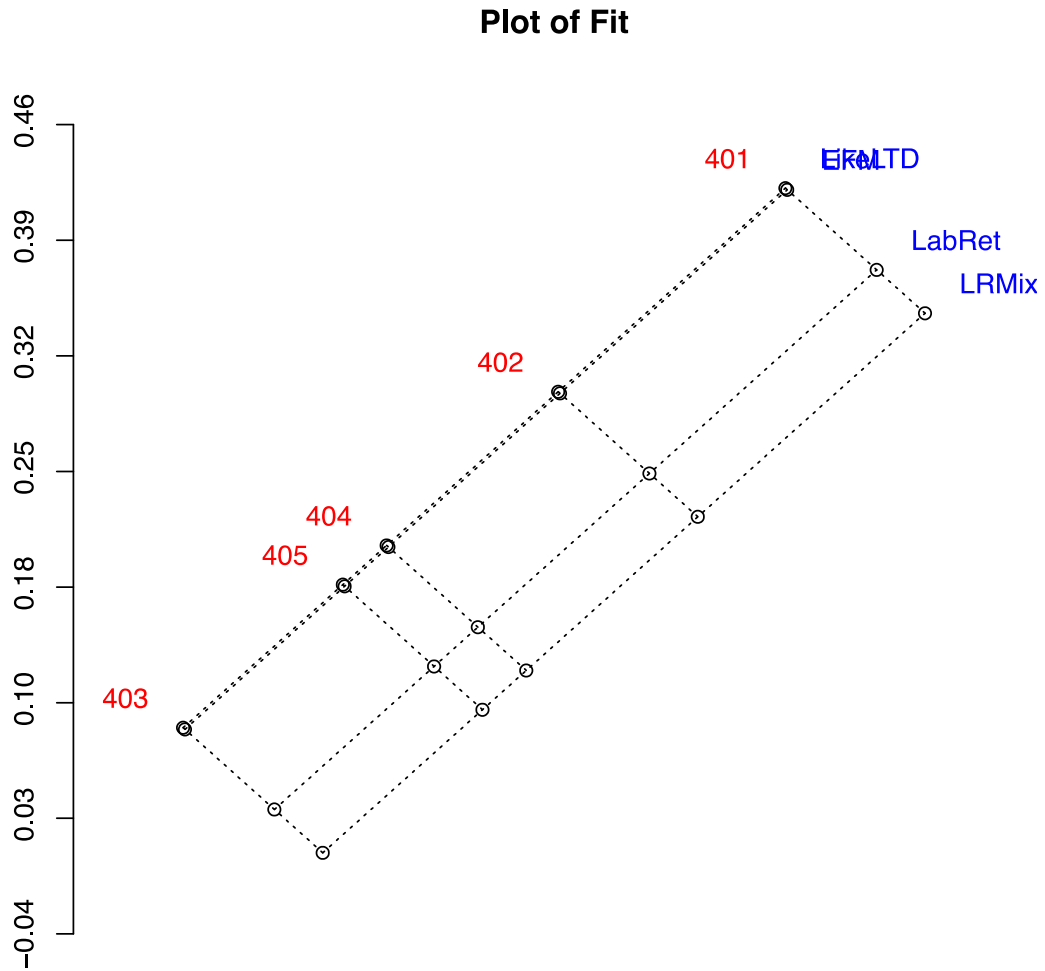
30pg 2:1


401-405
Replicate Variation
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
LabRet
likeLTD
EFM
LRMix

30pg 2:1

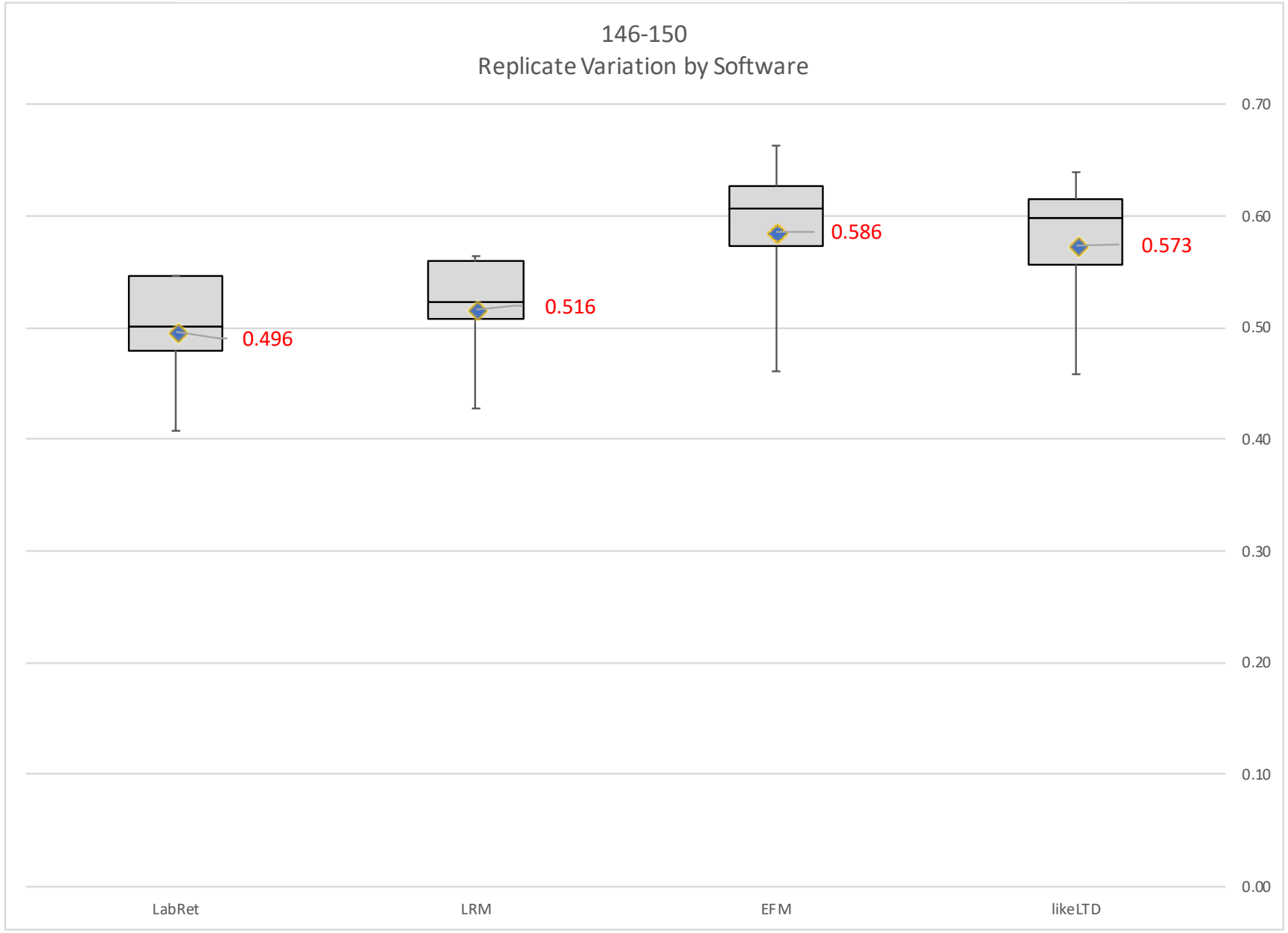

Replicates 401-405
Software Variation
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5

Median polish

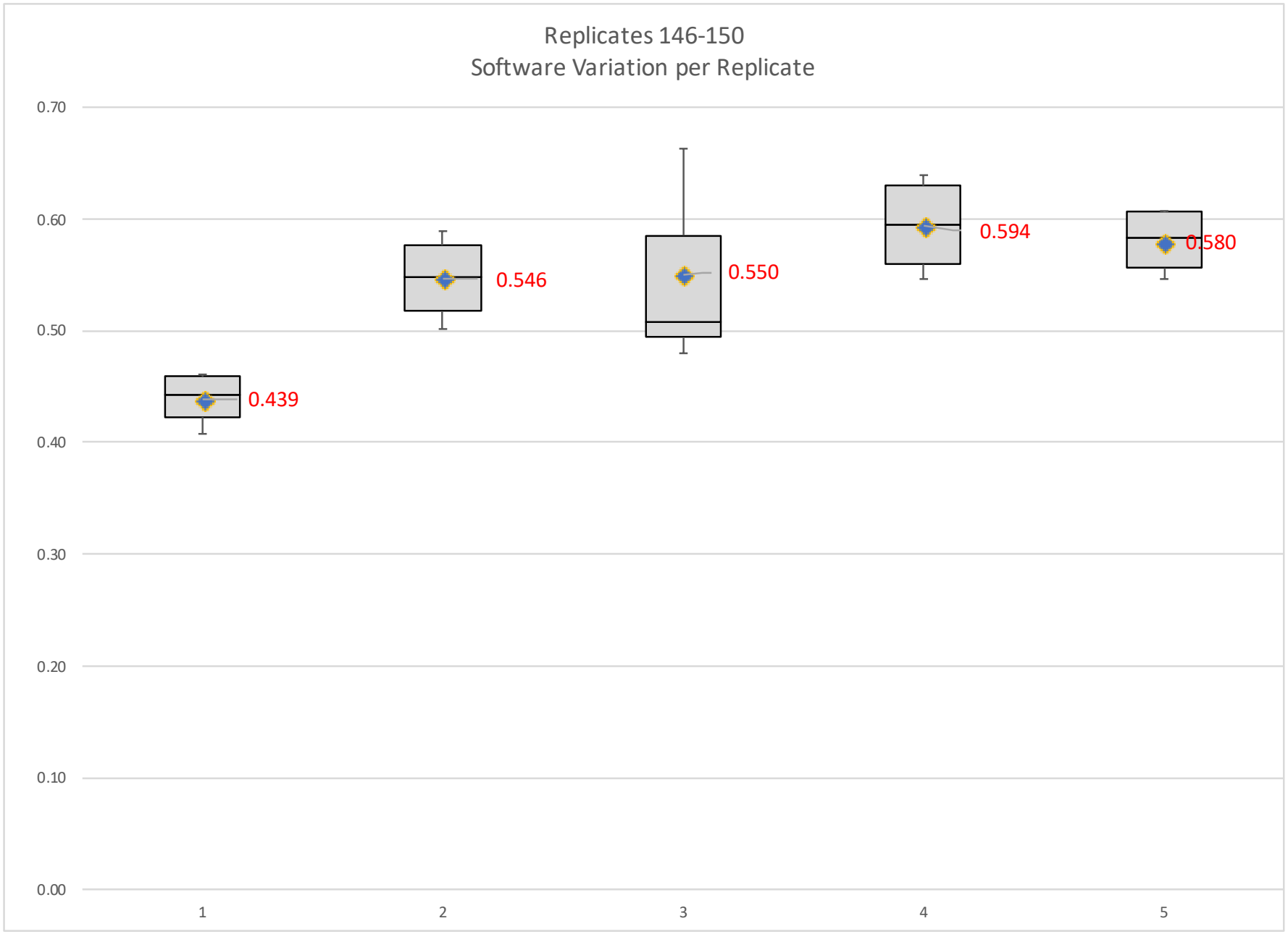


SDreps/SDalgs 3.3275

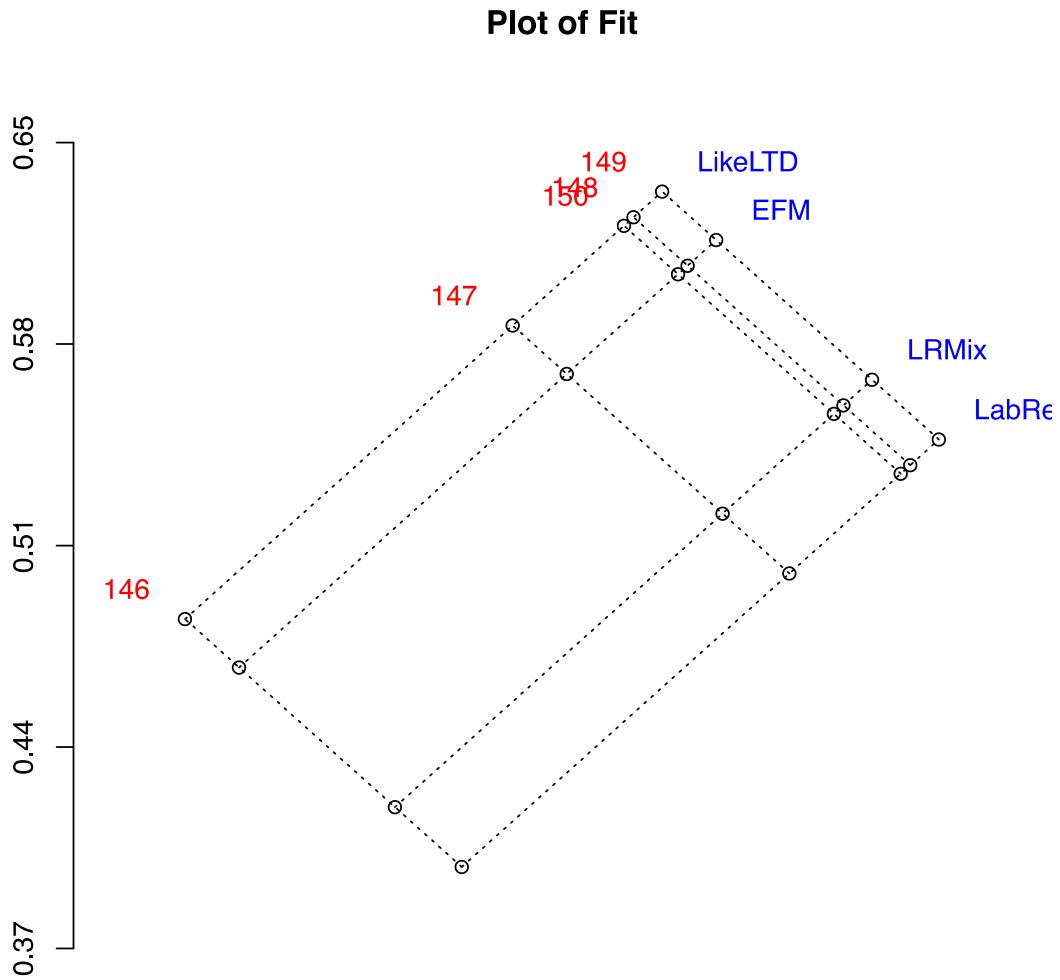
100pg
1:1


146-150
Replicate Variation by Software
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.496
0.516
0.586
0.573
LabRet
LRM
EFM
likeLTD

100pg
1:1


Replicates 146-150
Software Variation per Replicate
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.439
0.546
0.550
0.594
0.580
1
2
3
4
5

Median polish



The tentative summary

- Thresholds make a difference
  - We need all of the data present to make an informed inference from our mathematical models
- Replicates make a difference
  - We don’t know yet how many replicates are needed to capture all of the data in the extraction tube
    - Three is definitely not enough at the margins
- Not clear yet how much DNA or ratio of contributors defines the margins

THIS IS FOR TWO CONTRIBUTORS!

- Can we/should we consider replicates under most circumstances?
  - A Bayesian network would be a dandy tool to have

This work was supported by

- NIJ grant 2013-DN-BX-K029
- UK Engineering and Physical Sciences Research Council grant no EP/K032208/1 at the Isaac Newton Institute for Mathematical Sciences
- The Leverhulme Research Centre for Forensic Science