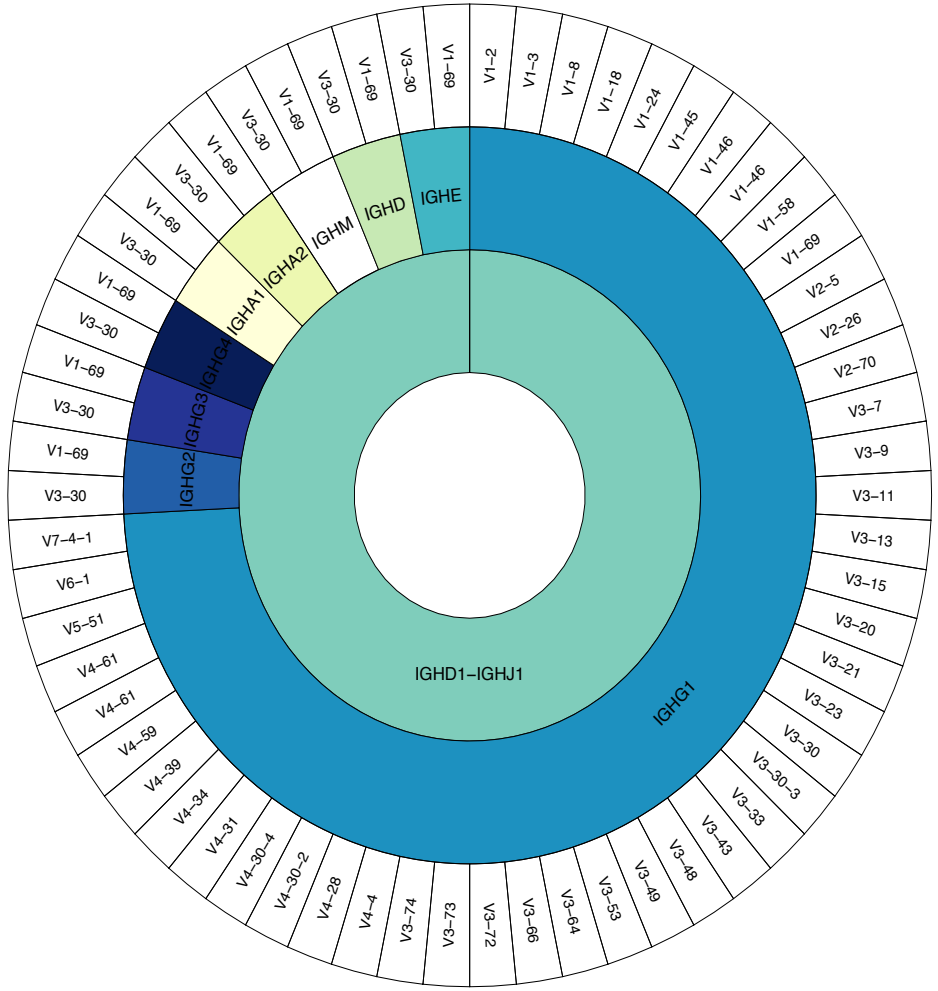
Development of synthetic standards for error correction in B-cell receptor repertoire sequencing

Johannes Trück

University Children's Hospital Zurich, Switzerland

Synthetic library

- 62 synthetic sequences
- 44 V-genes combined with all isotypes
- Random nt at VD and JC junctions
- Identical sequence length: 511 nt
- Synthetic tag between D and J: 18 nt / 6 aa (GALSQN)
- Pooled at same concentrations on cDNA level



Assessment of synthetic libraries

6 synthetic libraries amplified using **multiplex** primer mix (V family + MD/AEG)

- 10‘000 input cDNA molecules
- 50‘000
- 100‘000 (3 independent samples)
- 500‘000

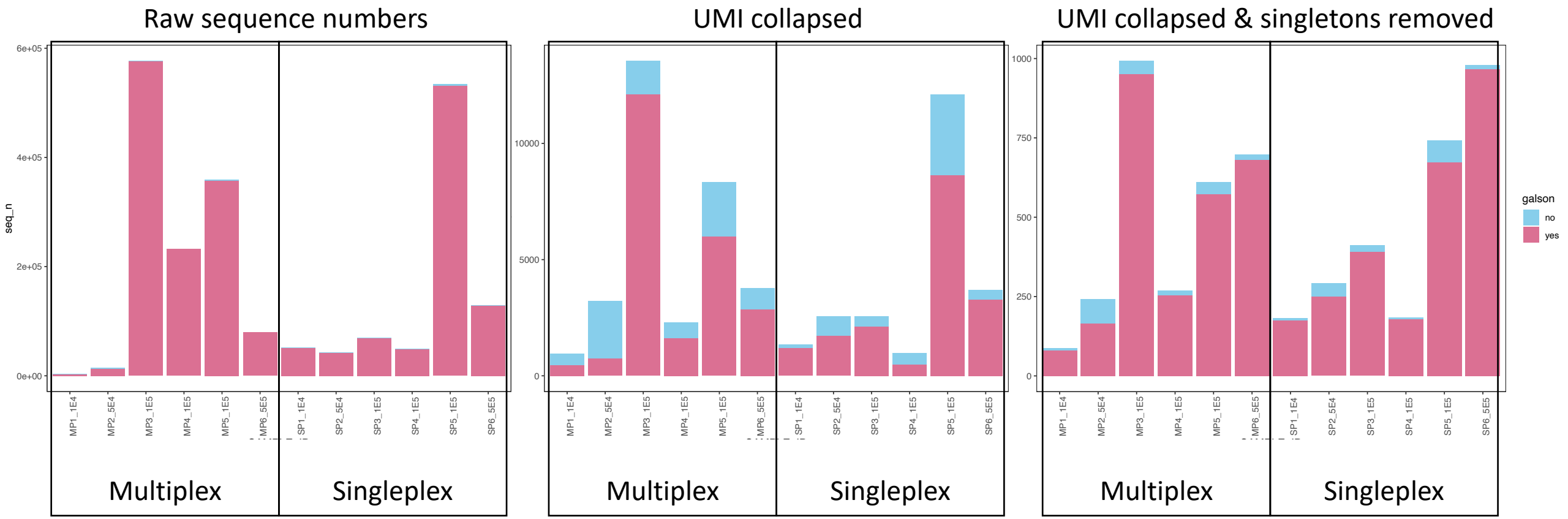
6 synthetic libraries amplified using **singleplex** primers (+ MD/AEG)

- 10‘000 input cDNA molecules
- 50‘000
- 100‘000 (3 independent samples)
- 500‘000

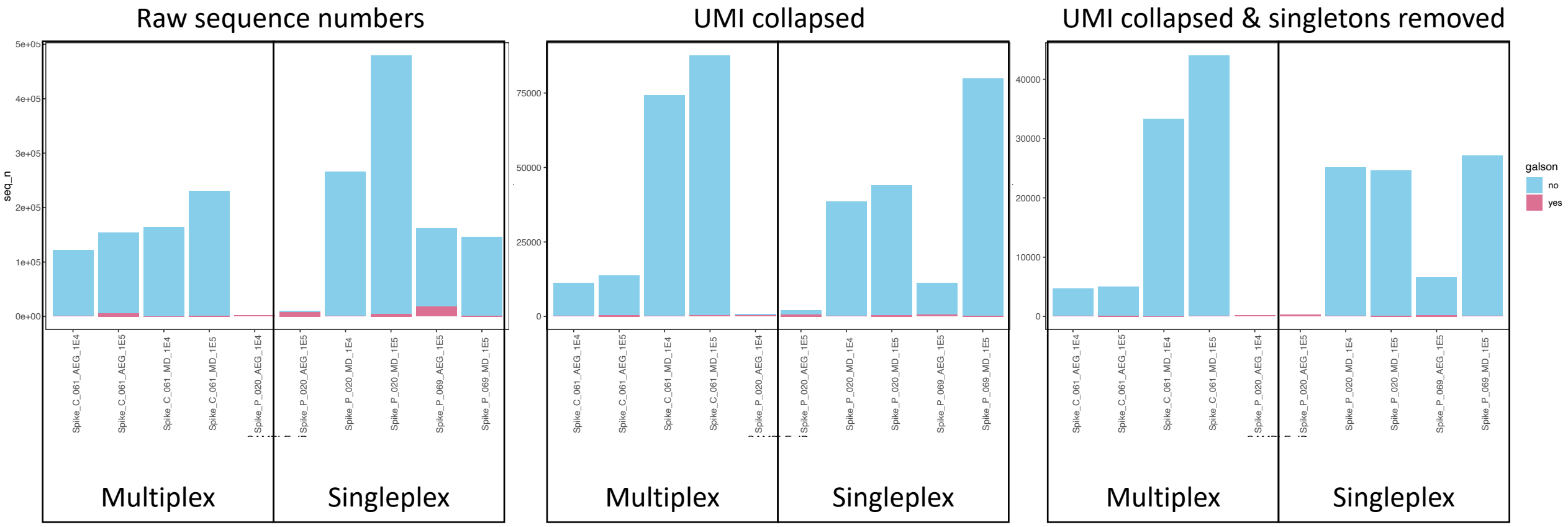
5 synthetic **spike-ins** (MD and AEG separately – biological cDNA conc. not measured)

- C.061 (10‘000 & 100‘000 input molecules)
- P.020 (10‘000 & 100‘000 input molecules)
- P.069 (100‘000 input molecules)

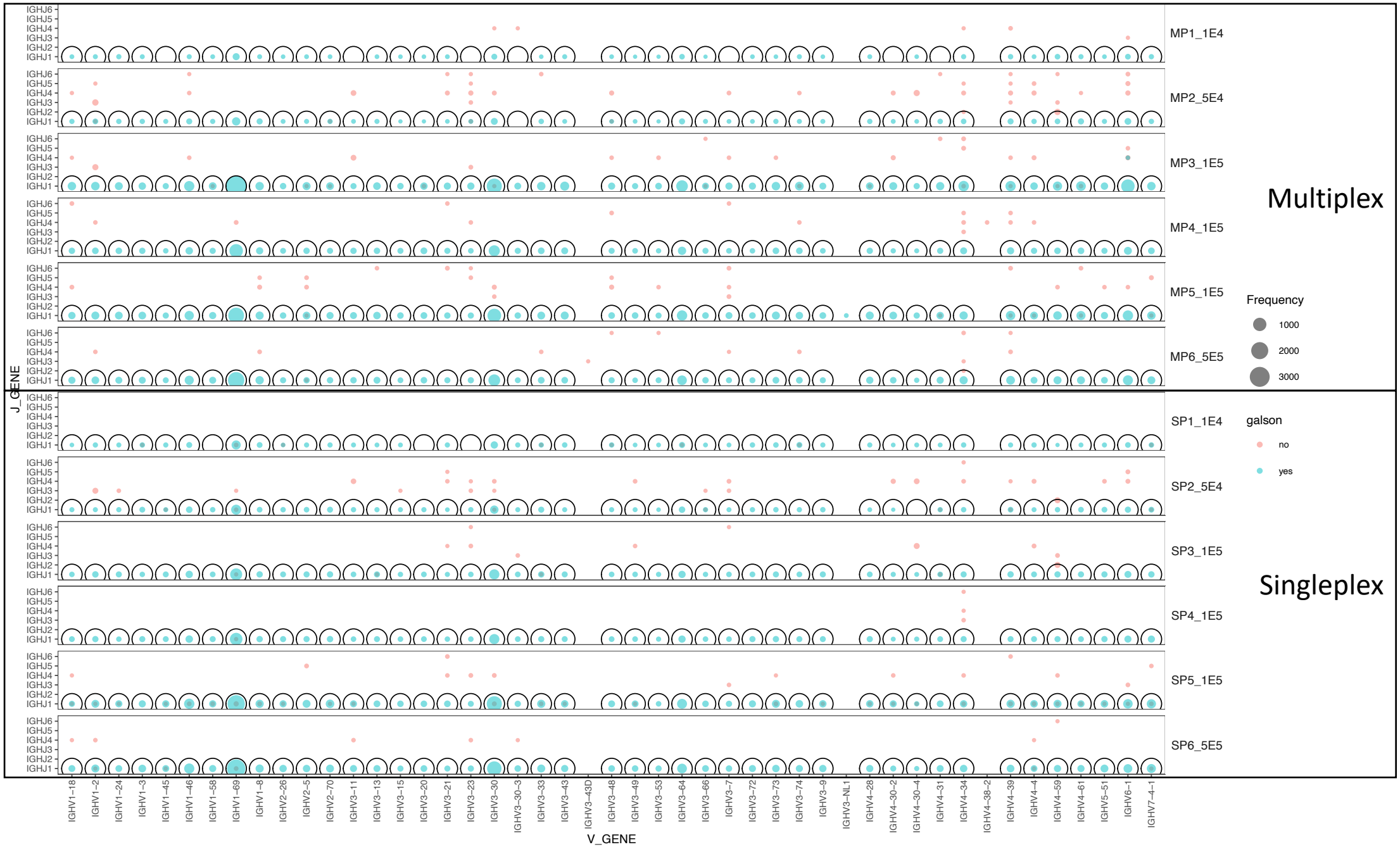
Sequence numbers – **Synthetic samples**



Sequence numbers – **Spike-in samples**



Sensitivity – **Synthetic samples**

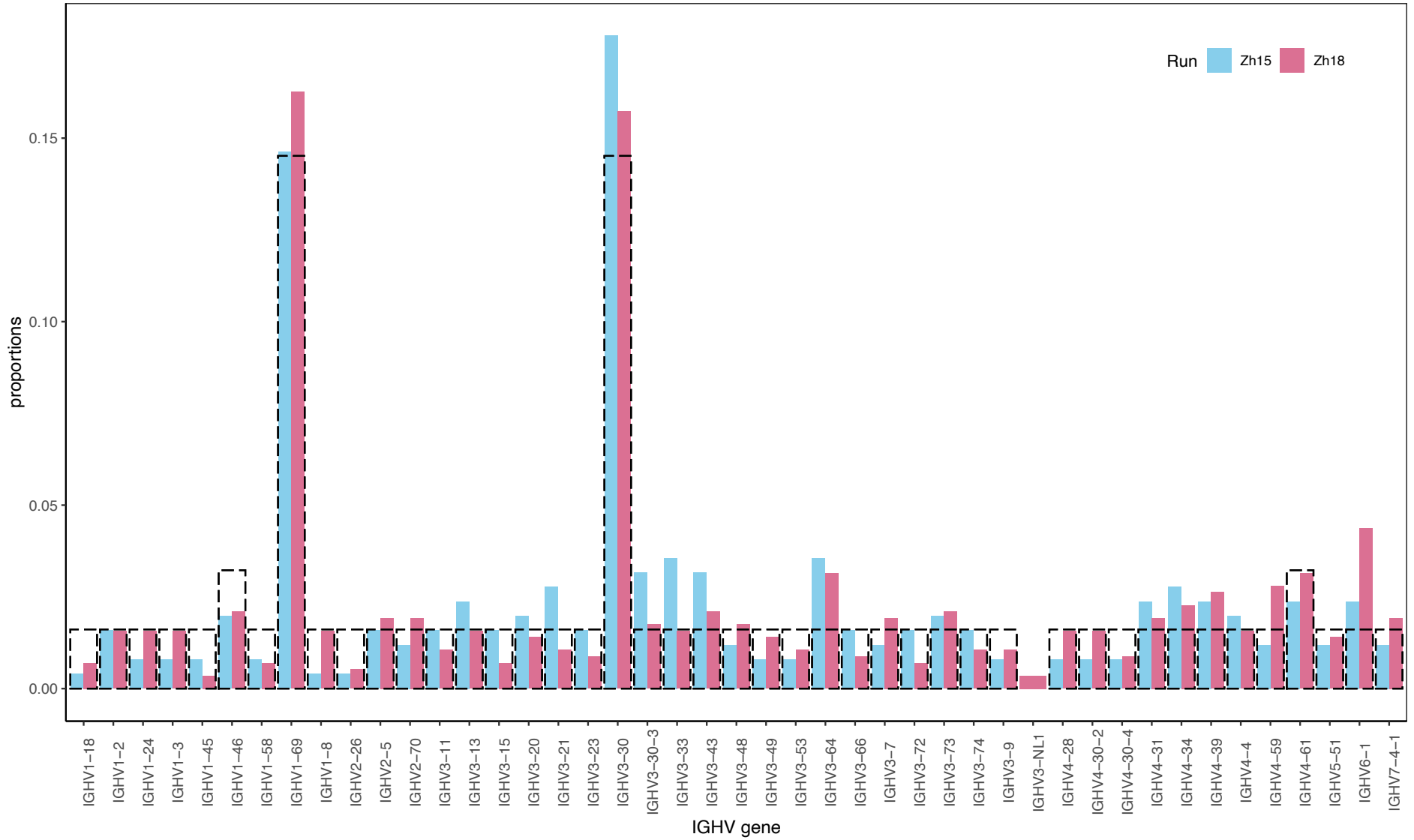


concentration

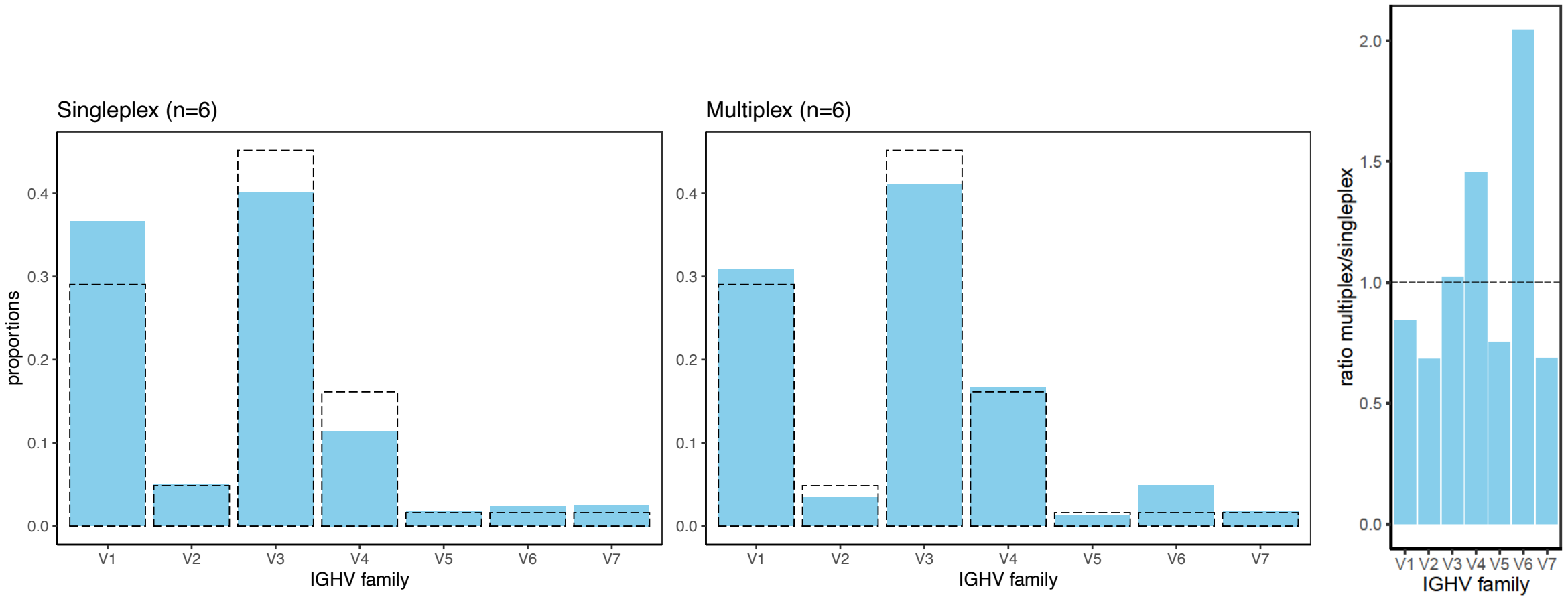
concentration

Sensitivity – **Spike-in samples**

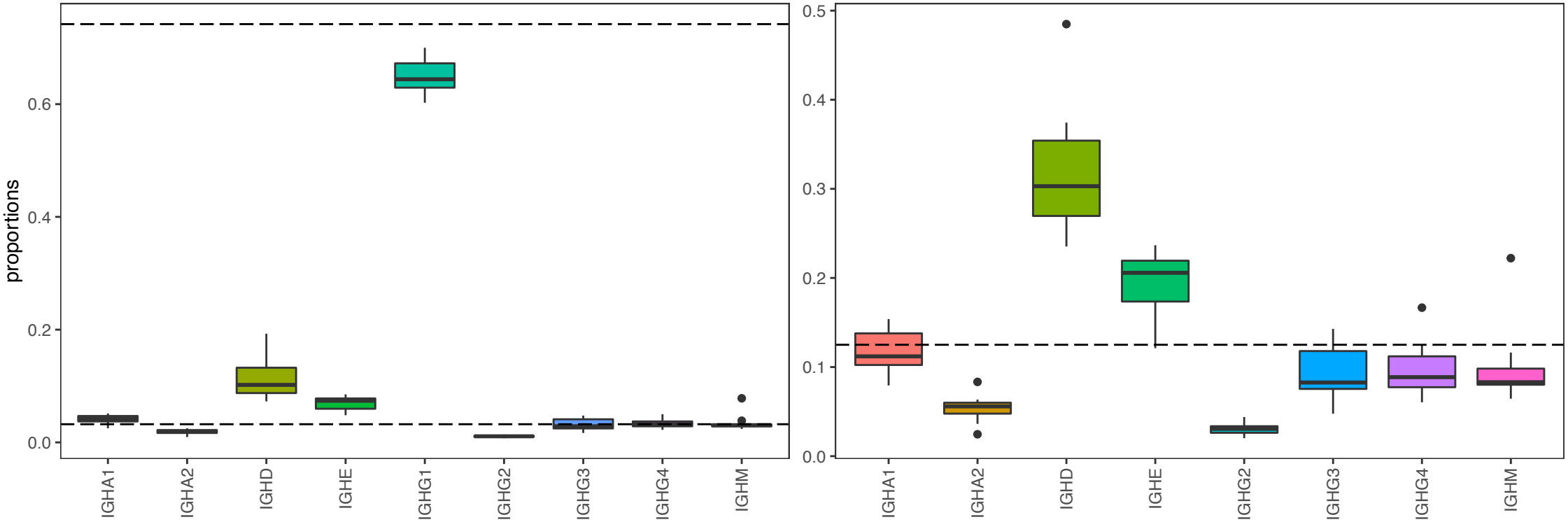
Multiplex PCR – V gene frequencies



Multiplex vs. singleplex PCR – **V family frequencies**



Isotype subclass frequencies



Summary

- Successful development of a synthetic library with n = 62 sequences
  - Identical length
  - 18 nt tag within CDR3
  - Pooled at same concentration
- Helpful to assess current PCR amplification method
  - Some systematic bias but no “hole”
  - Identification of “biological” sequences in synthetic samples
- Synthetic spike-ins challenging