

Preparing for the data deluge

E-Infrastructure *challenges* for the life sciences

EBI • Hinxton • May 2014

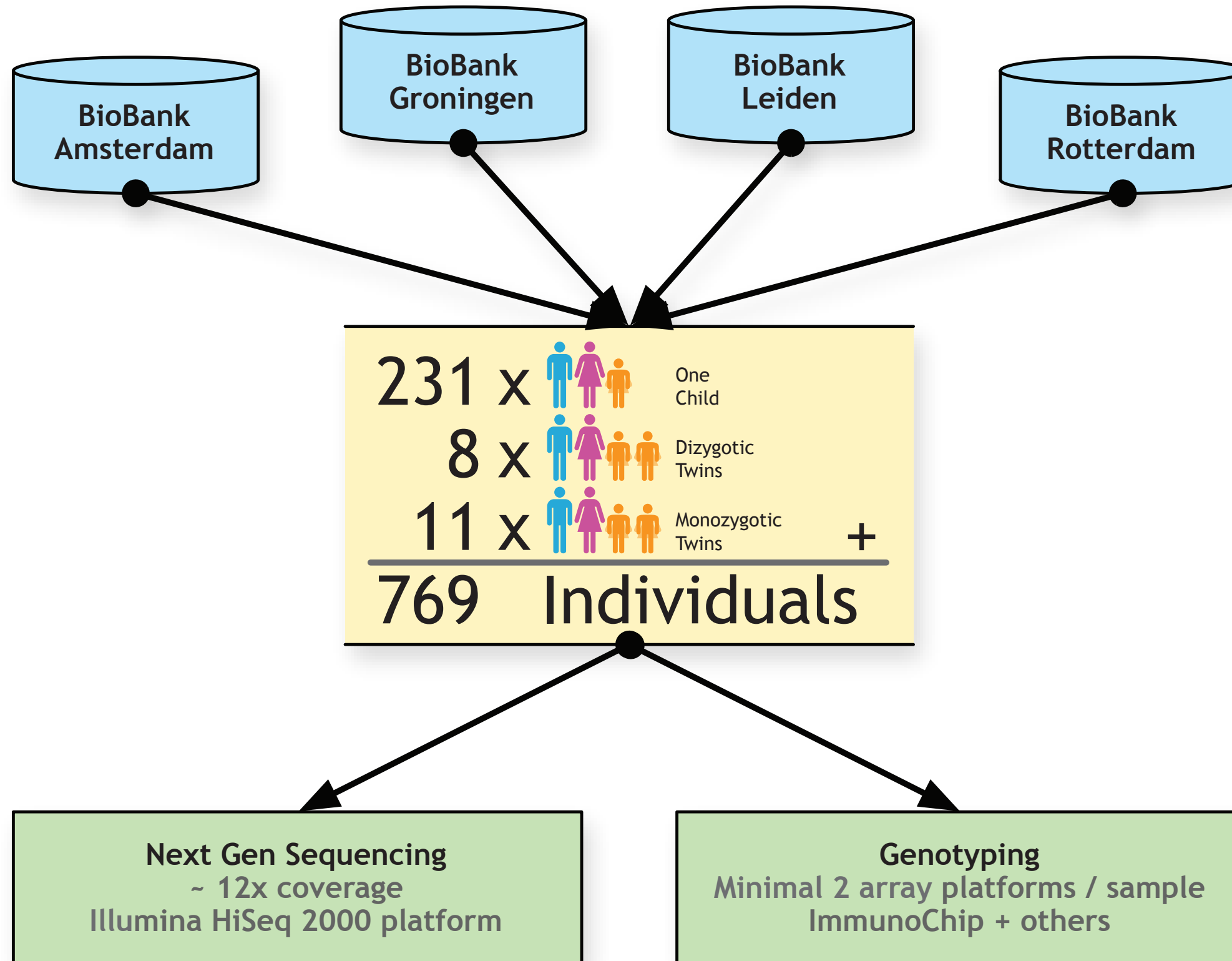


university of
 groningen



umcg

GoNL



GoNL

GoNL
GENOMEoftheNETHERLANDS



- Research
 - BBMRI • GoNL
 - BBMRI • BIOS
 - CTMM • TraIT
- Diagnostics
 - 5GPM
- e-Infra
 - Clusters
 - Grid
 - Partially IO bound

Challenges

- Managing data from Biobanks / Hospitals
 - Big data: unique possibilities for personalised medicine
 - Access and Privacy
 - DNA profile != Anonymous
 - Possible solutions
 - Share only summaries / groups
 - Distributed queries

Challenges

- Managing data
 - QC of large data sets
 - Effect of substantial artefact may be diluted in pool of correct measurements
 - Summary stats not sufficient
 - Manual inspection not feasible
 - eLabjournal
 - Analysis (jobscrip) logs capture how analysis was performed,
 - but do not log *why* analysis was performed / repeated...

Challenges

- Managing data
 - Distribution: disks per mail and 10 Gbit network

