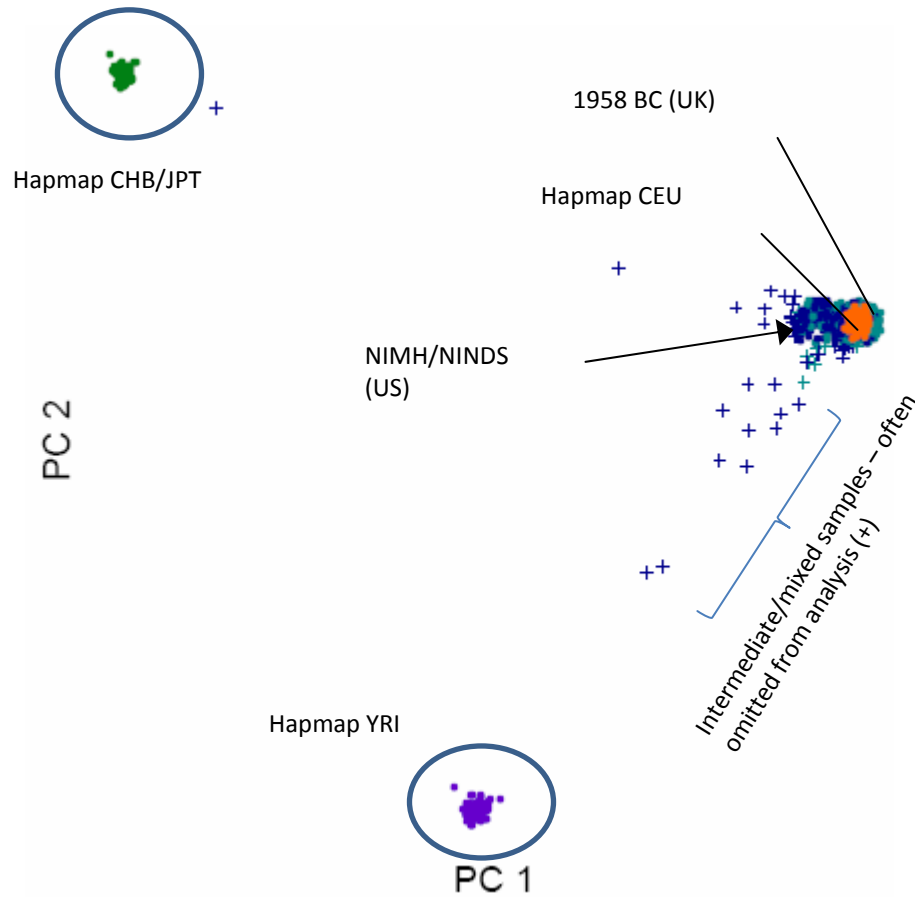


Minimal information for databases

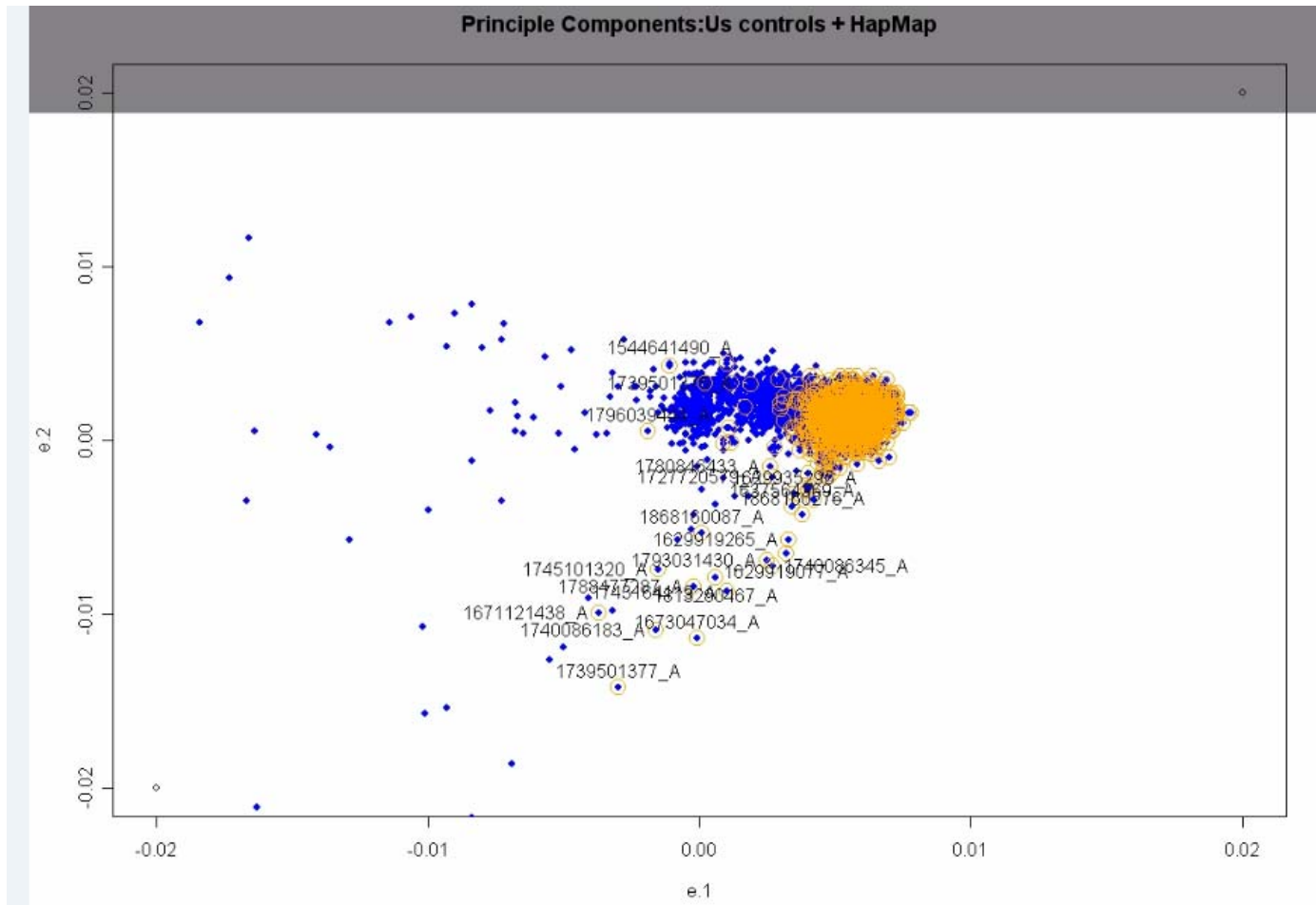
Minimal information

- Minimal information involves ‘raw’ data, not just summaries
- ‘Raw’ genotype data has several definitions
 - Genotypes change, primary data do not
- It is *very* easy to get it wrong
 - Strand differences
 - Allele calling: several algorithms, dynamic
 - Automated QC doesn’t catch everything (even simple ‘phenotypes’)
 - New uses of clean data reveal new problems

Matching on genetic background



Diversity in shared data



Needs

- As much 'raw' data as absolutely possible
 - Many as-yet undefined uses
 - Different uses will reveal different features, errors
- Need iterative annotation mechanism
 - Individual's DNA is fixed, but (measured) genotype is not
 - Many people working hard on public genotypes – need way to feed this information back into dbGaP
- Making data available is great first step, but collaboration is even better