International Statistical Genetics Workshop
IBG
Boulder, Colorado

Quality Control

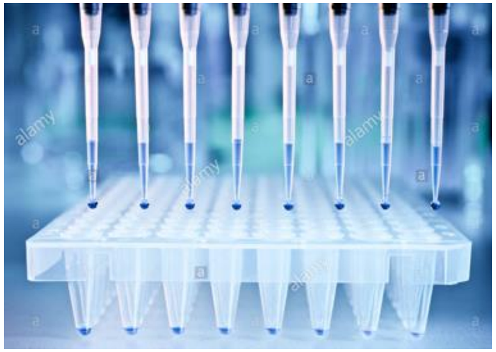
Katrina Grasby and Lucia Colodro Conde

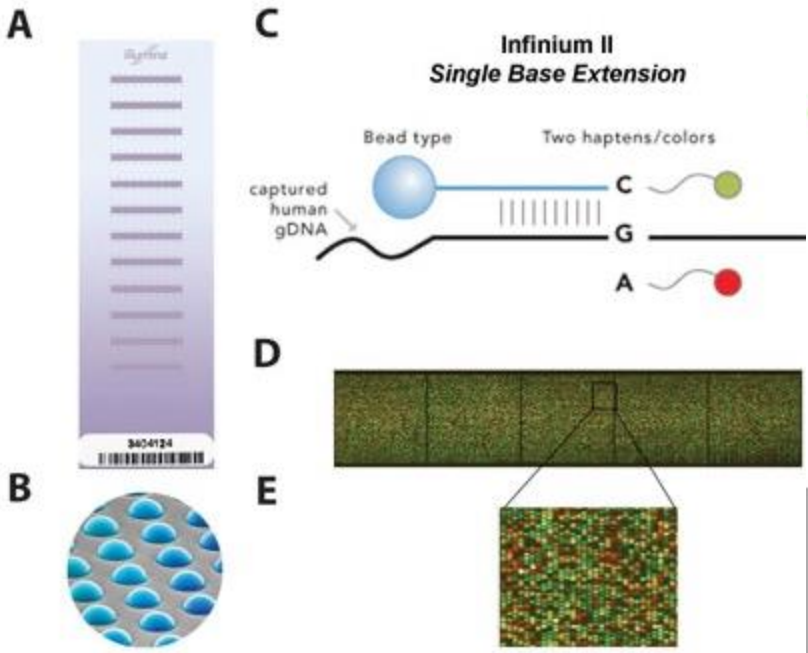
Why Do Quality Control?

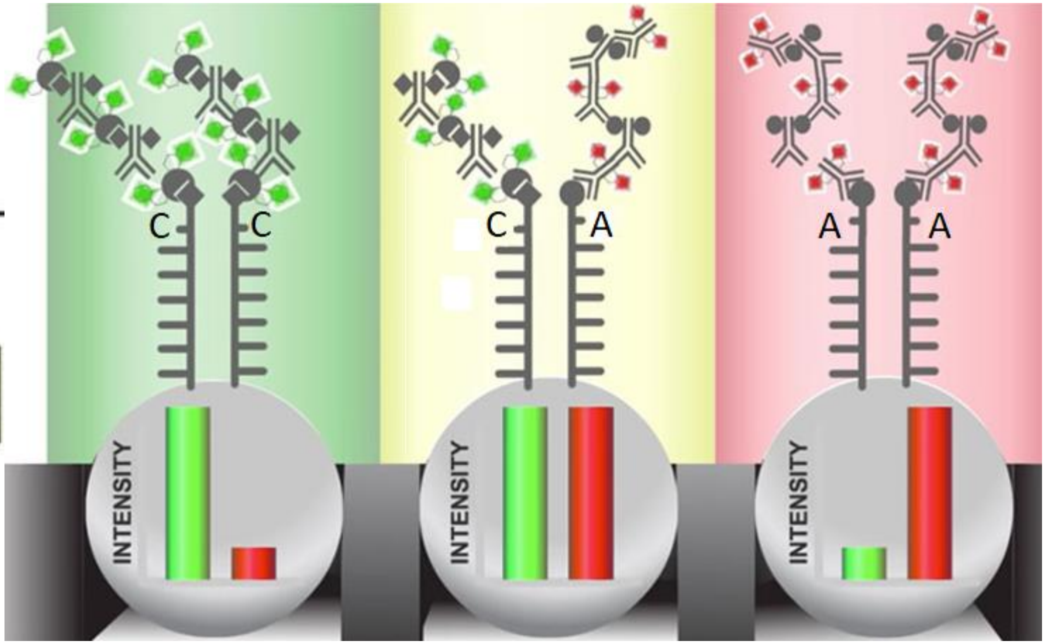
Poor quality data → false positives / negatives

- To remove genotyping errors
    - Low quality or quantity of DNA
    - Contaminated DNA
    - Chemical or machinery failure
    - Human error
- To ensure data suitable for the analyses
    - Relatedness

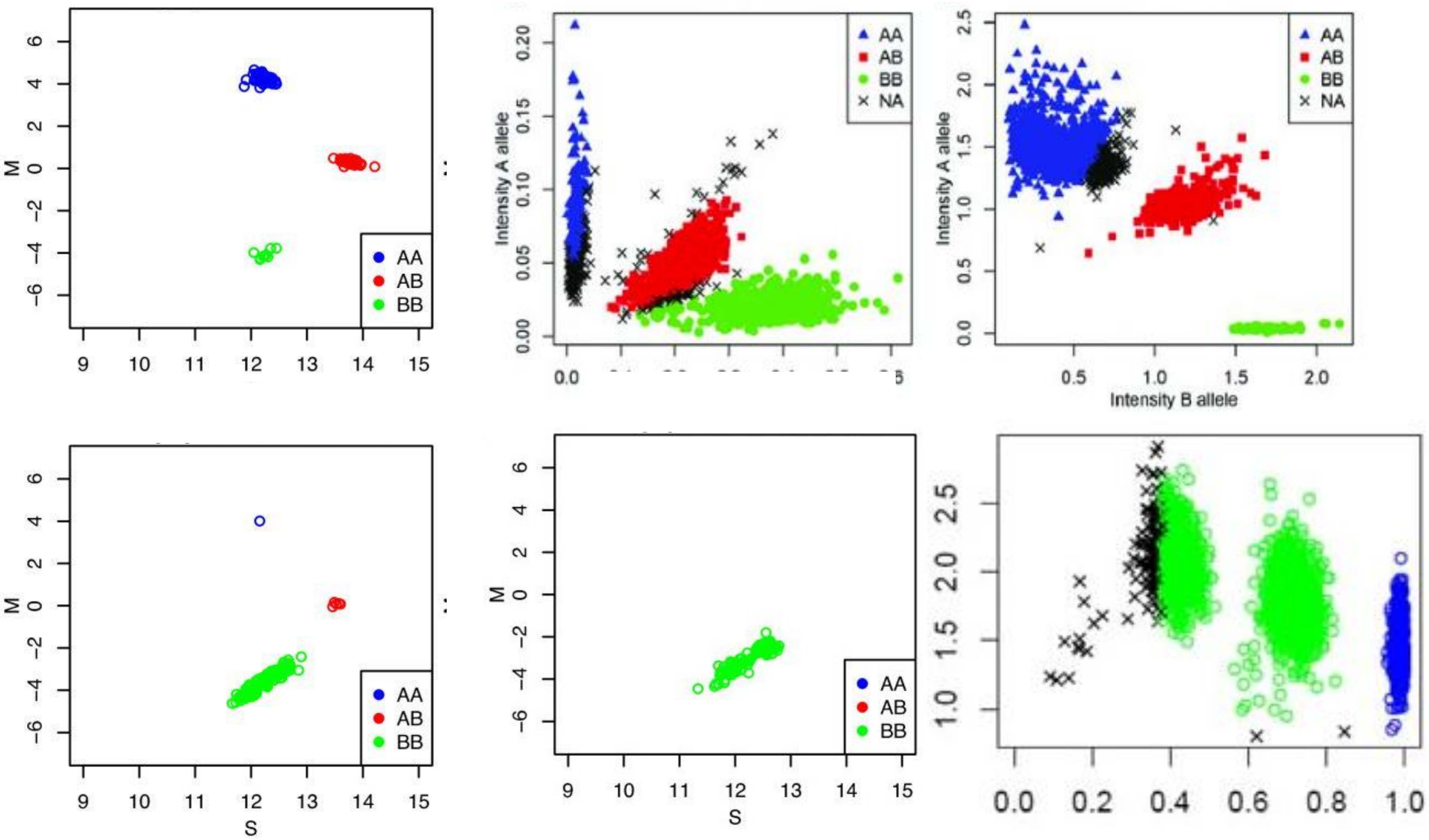
From DNA to data







Genotyping Intensities



Key Steps in QC

1. Check Data (file format, data coding, missing, build)

Check the build

https://genome.ucsc.edu

Human Assembly

- ✓ Dec. 2013 (GRCh38/hg38)
- Feb. 2009 (GRCh37/hg19)
- Mar. 2006 (NCBI36/hg18)
- May 2004 (NCBI35/hg17)

Key Steps in QC

1. Check Data (file format, data coding, missing, build)
2. Sex (chr X heterozygosity)

Key Steps in QC

1. Check Data (file format, data coding, missing, build)
2. Sex (chr X heterozygosity)
3. Genotyping Call Rate (SNPs missing individuals)

Key Steps in QC

1. Check Data (file format, data coding, missing, build)
2. Sex (chr X heterozygosity)
3. Genotyping Call Rate (SNPs missing individuals)
4. Hardy-Weinberg Equilibrium

Key Steps in QC

1. Check Data (file format, data coding, missing, build)
2. Sex (chr X heterozygosity)
3. Genotyping Call Rate (SNPs missing individuals)
4. Hardy-Weinberg Equilibrium
5. Minor Allele Frequency

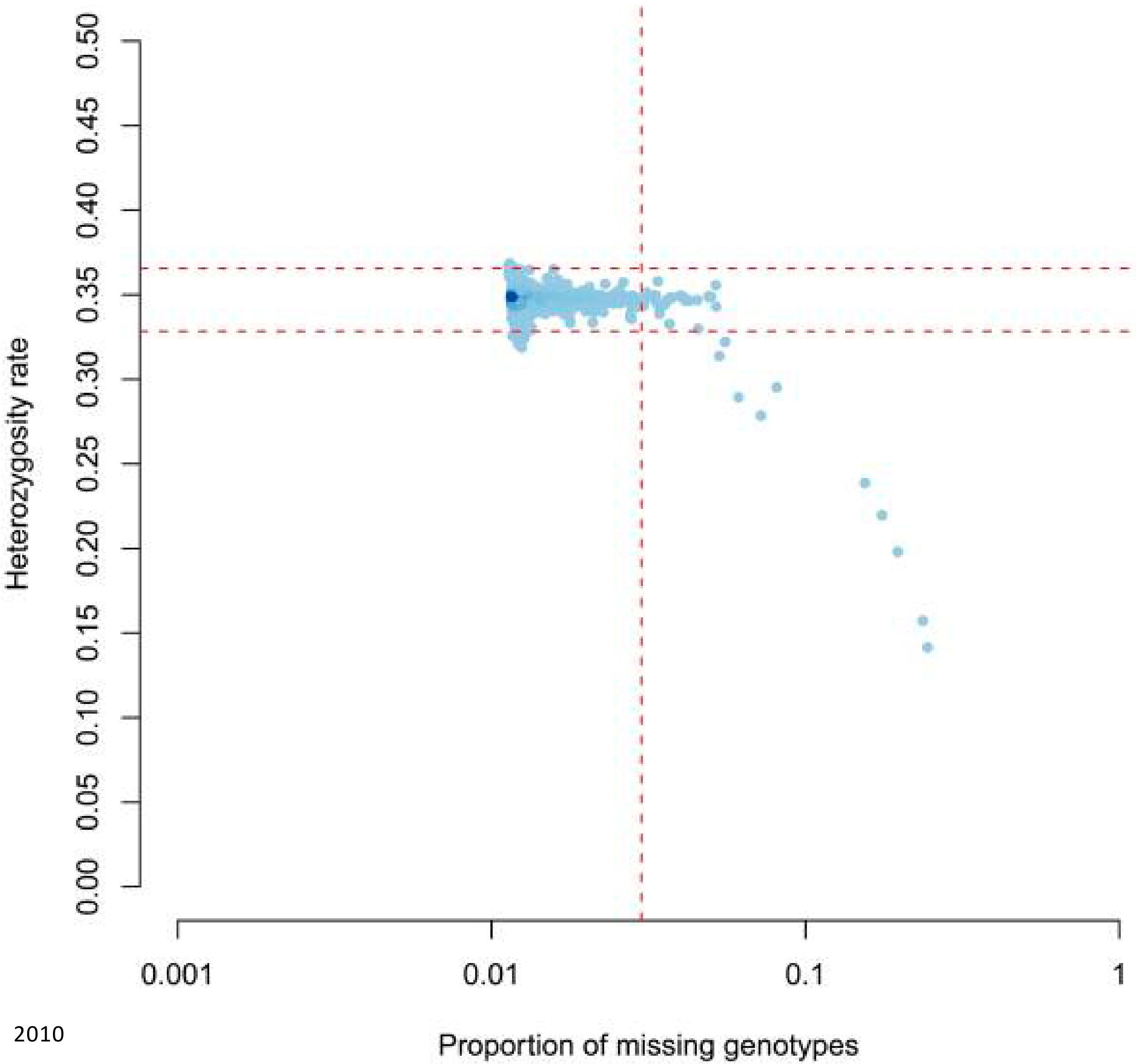
Key Steps in QC

1. Check Data (file format, data coding, missing, build)
2. Sex (chr X heterozygosity)
3. Genotyping Call Rate (SNPs missing individuals)
4. Hardy-Weinberg Equilibrium
5. Minor Allele Frequency
6. Sample Call Rate (individuals missing genotypes)

Key Steps in QC

1. Check Data (file format, data coding, missing, build)
2. Sex (chr X heterozygosity)
3. Genotyping Call Rate (SNPs missing individuals)
4. Hardy-Weinberg Equilibrium
5. Minor Allele Frequency
6. Sample Call Rate (individuals missing genotypes)
7. Proportion of Heterozygosity

Heterozygosity vs. Missing



Anderson et al. *Nat Proc.* 2010

Key Steps in QC

1. Check Data (file format, data coding, missing, build)
2. Sex (chr X heterozygosity)
3. Genotyping Call Rate (SNPs missing individuals)
4. Hardy-Weinberg Equilibrium
5. Minor Allele Frequency
6. Sample Call Rate (individuals missing genotypes)
7. Proportion of Heterozygosity
8. Relatedness

Key Steps in QC

1. Check Data (file format, data coding, missing, build)
2. Sex-check (chr X heterozygosity)
3. Genotyping Call Rate (SNPs missing individuals)
4. Hardy-Weinberg Equilibrium
5. Minor Allele Frequency
6. Sample Call Rate (individuals missing genotypes)
7. Proportion of Heterozygosity
8. Relatedness
9. Population Structure / Stratification

Key Steps in QC

1. Check Data (file format, data coding, missing, build)
2. Sex-check (chr X heterozygosity)
3. Genotyping Call Rate (SNPs missing individuals)
4. Hardy-Weinberg Equilibrium
5. Minor Allele Frequency
6. Sample Call Rate (individuals missing genotypes)
7. Proportion of Heterozygosity
8. Relatedness
9. Population Structure / Stratification