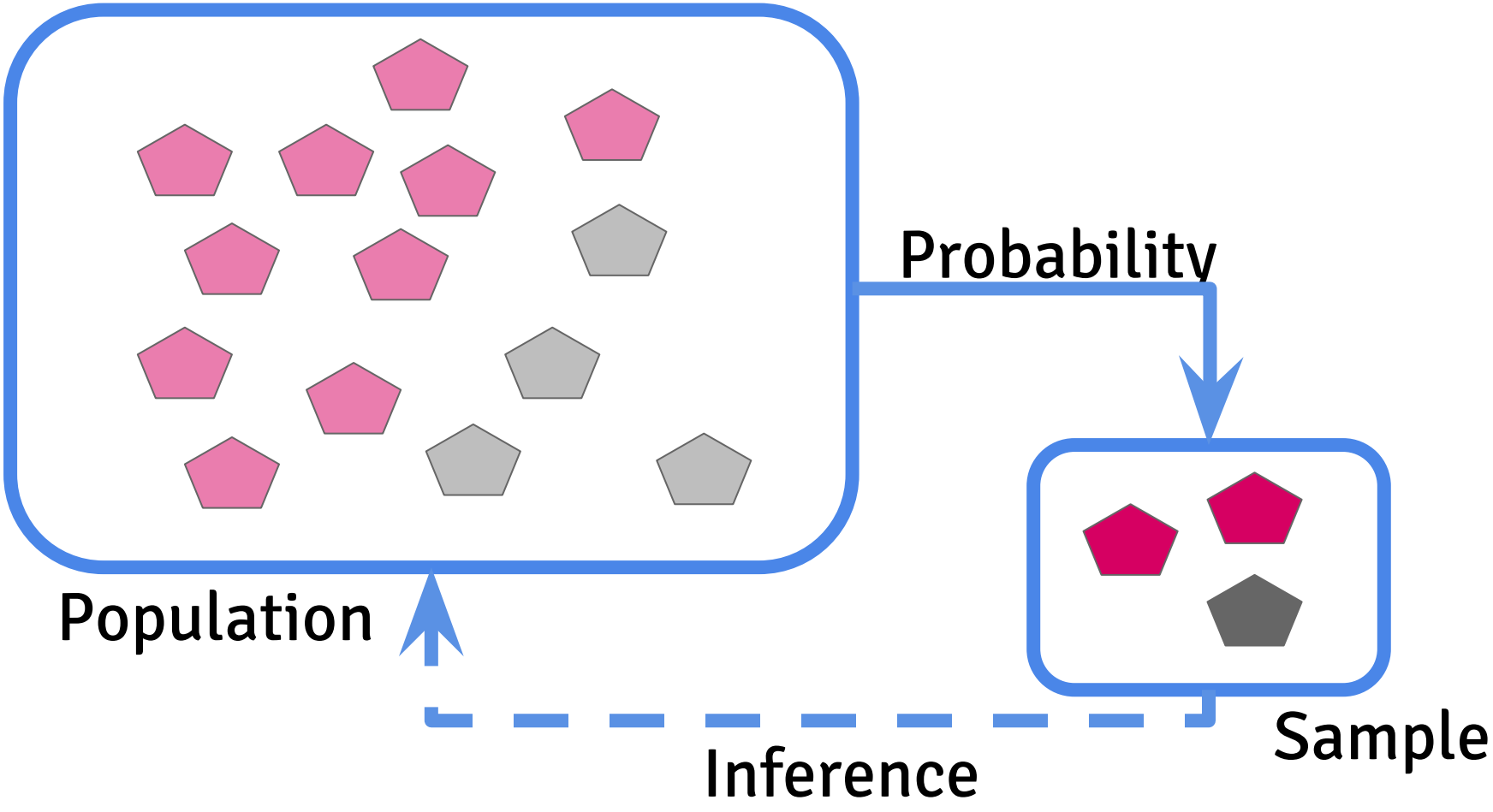
Sample size and variability

Jeff Leek

@jtleek

Central dogma of statistics

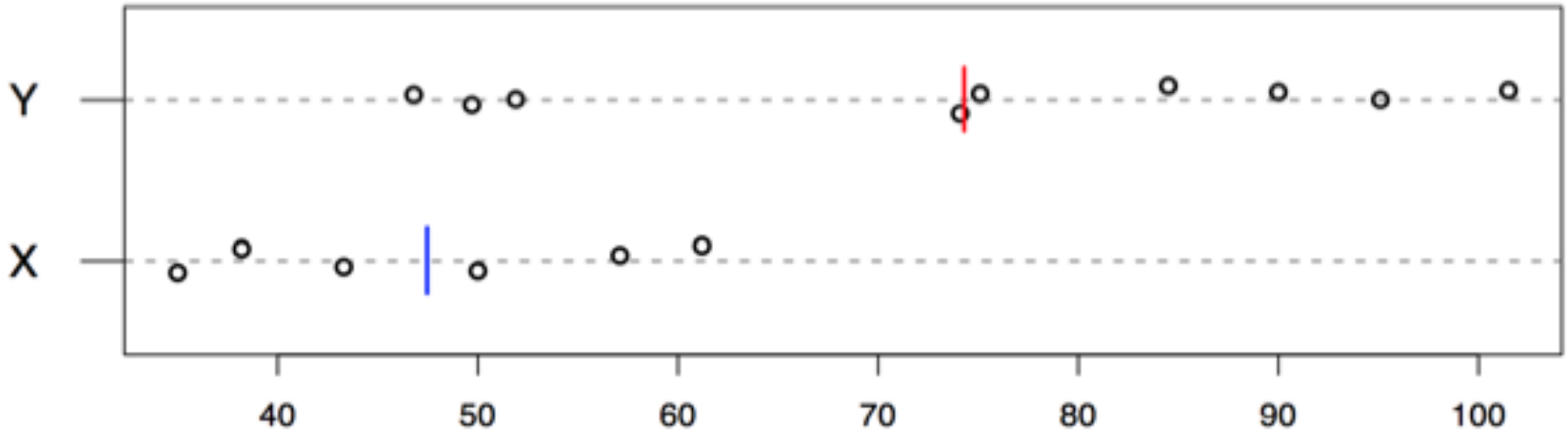

Probability
Population
Sample
Inference

Sample size

N = Number of Measurements

$$N = \frac{(\$ \text{ you have})}{(\$ \text{ / measurement})}$$

Variability and power


Y
X
40
50
60
70
80
90
100

- n = 10 for each group; effect = $\Delta = 5$; pop'n SD = $\sigma = 10$

  ```
  power.t.test(n=10, delta=5, sd=10)
  ```

  $\longrightarrow$ 18%

- power = 80%; effect = $\Delta = 5$; pop'n SD = $\sigma = 10$

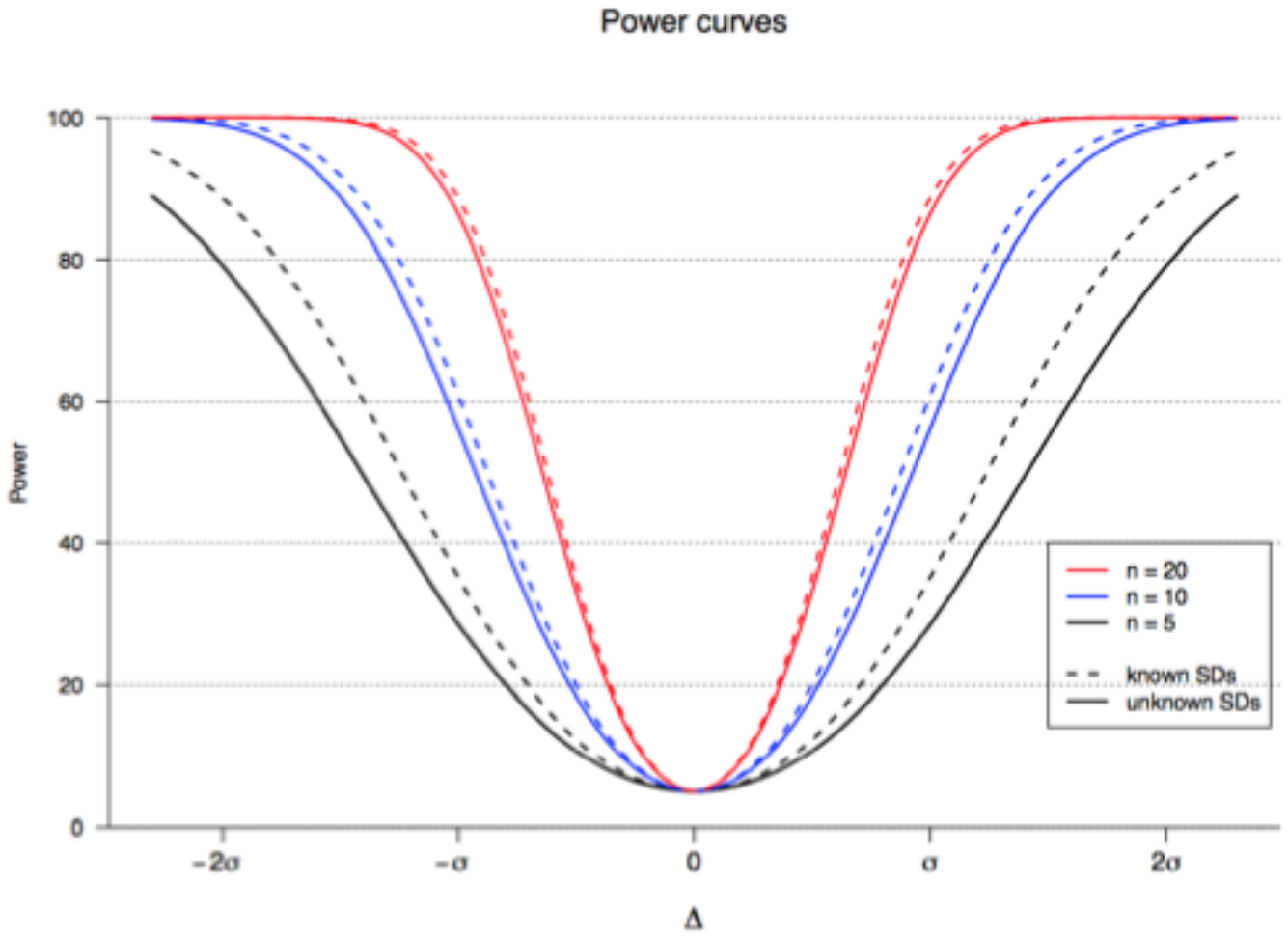
  ```
  power.t.test(delta=5, sd=10, power=0.8)
  ```

  $\longrightarrow$ n = 63.8 $\longrightarrow$ 64 for each group

- power = 80%; effect = $\Delta = 5$; pop'n SD = $\sigma = 10$; one-sided

  ```
  power.t.test(delta=5, sd=10, power=0.8,
               alternative="one.sided")
  ```

  $\longrightarrow$ n = 50.2 $\longrightarrow$ 51 for each group


Power curves
Power
100
80
60
40
20
0
−2σ
−σ
0
σ
2σ
Δ
n = 20
n = 10
n = 5
known SDs
unknown SDs

Three types of variablity

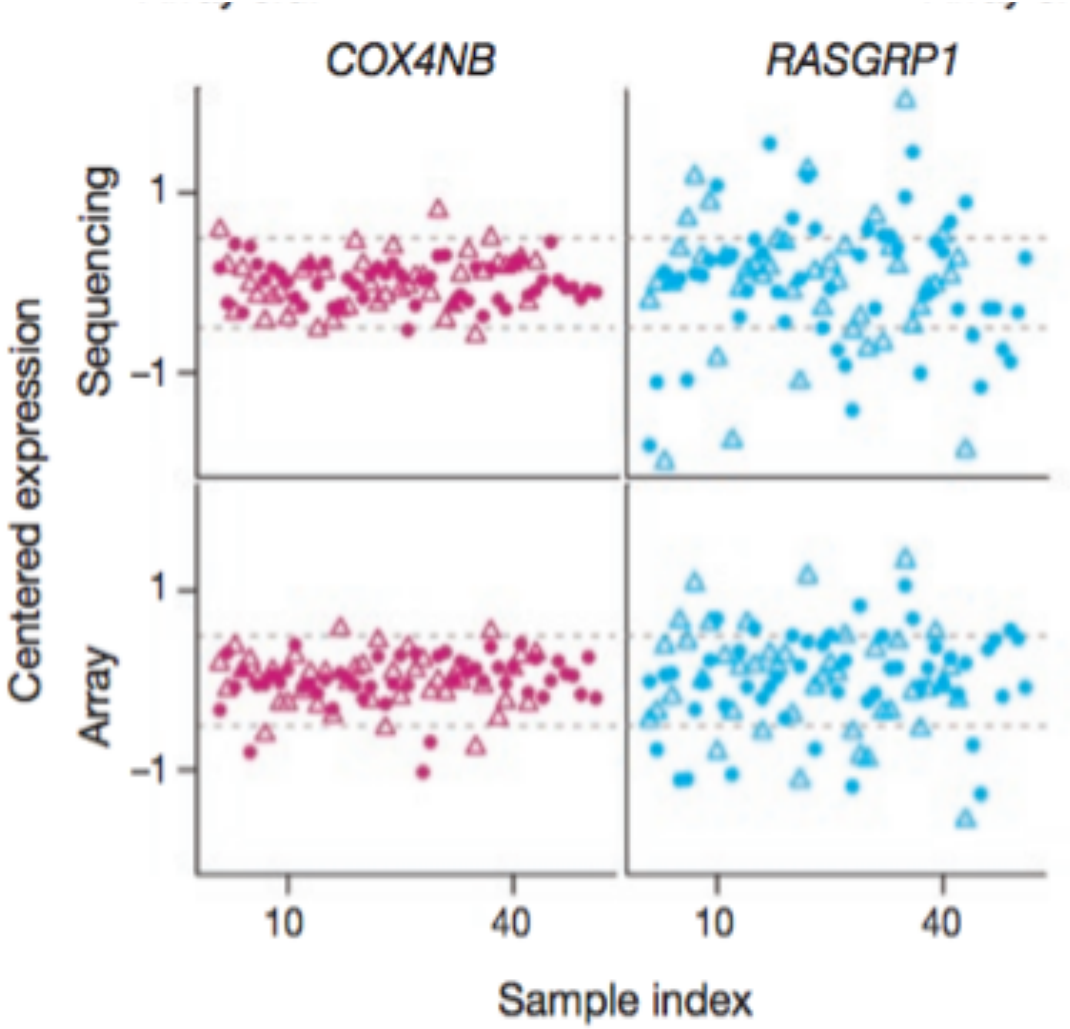
Var(Genomic Measurement)

= Phenotypic variability

+ Measurement error

+ Natural biological variation

New technology doesn’t eliminate variability


COX4NB
RASGRP1
Centered expression
Sequencing
Array
1
−1
1
−1
10
40
10
40
Sample index