Statistical Computing

Analysis of treatments that modify the DNA replication velocity and stability

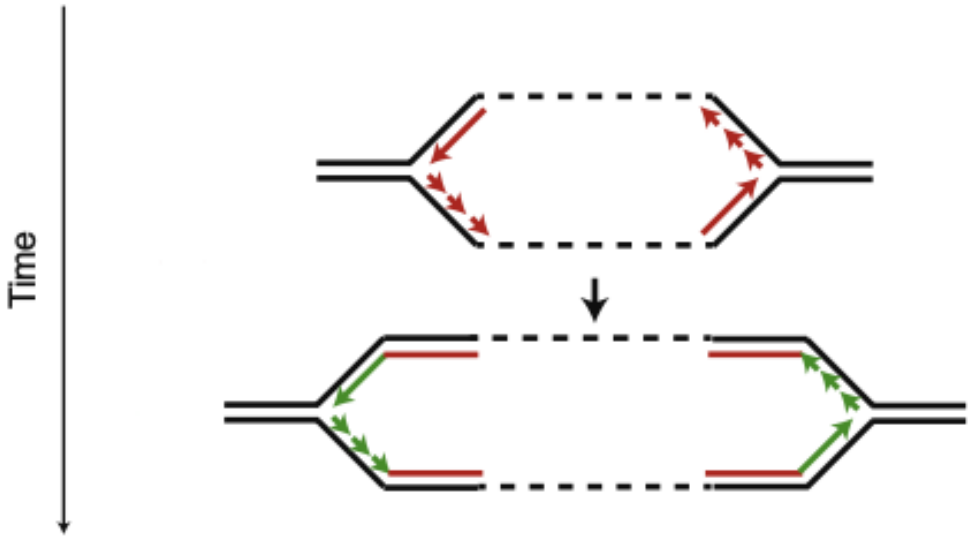
Uwe Menzel, 2017

uwe.menzel@matstat.org

www.matstat.org

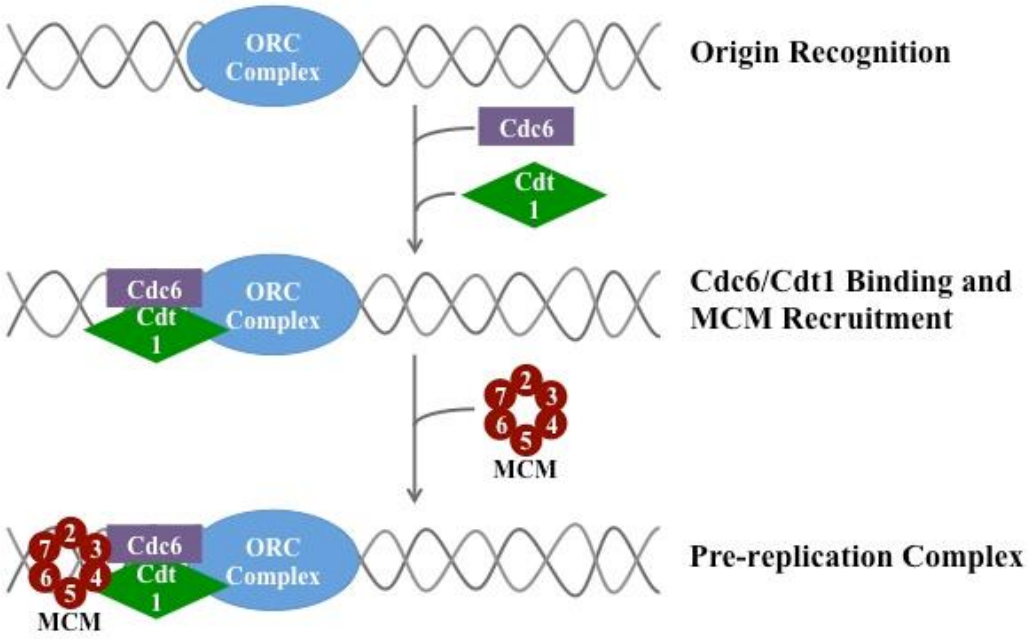
DNA replication

- produces two identical copies from one DNA molecule
- starts at replication origins
  - high AT content → easy unwinding of strands
  - thousands per chromosome, ~ 130 kB distance
- bidirectional ("fork") or unidirectional



DNA replication initiation (Eukaryotes)

- unwinding (helicases, DHX9)
- origin recognition complex binds to replication origin
- a number of proteins are being recruited (Cdc45)



→ DNA replication start

Papers

Cdc45 is limiting for replication initiation in humans

Carsten Köhler[a,†], Dennis Koalick[a,†], Anja Fabricius[a], Ann Christin Parplys[b], Kerstin Borgmann[b], Helmut Pospiech[a,c], and Frank Grosse[a,d]

[a]Research group Biochemistry, Leibniz Institute for Age Research – Fritz Lipmann Institute, Jena, Germany; [b]Laboratory of Radiobiology and Experimental Radiation Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [c]Faculty of Biochemistry and Molecular Medicine, University of Oulu, Finland; [d]Centre for Molecular Biomedicine, Friedrich-Schiller University, Jena, Germany

ABSTRACT

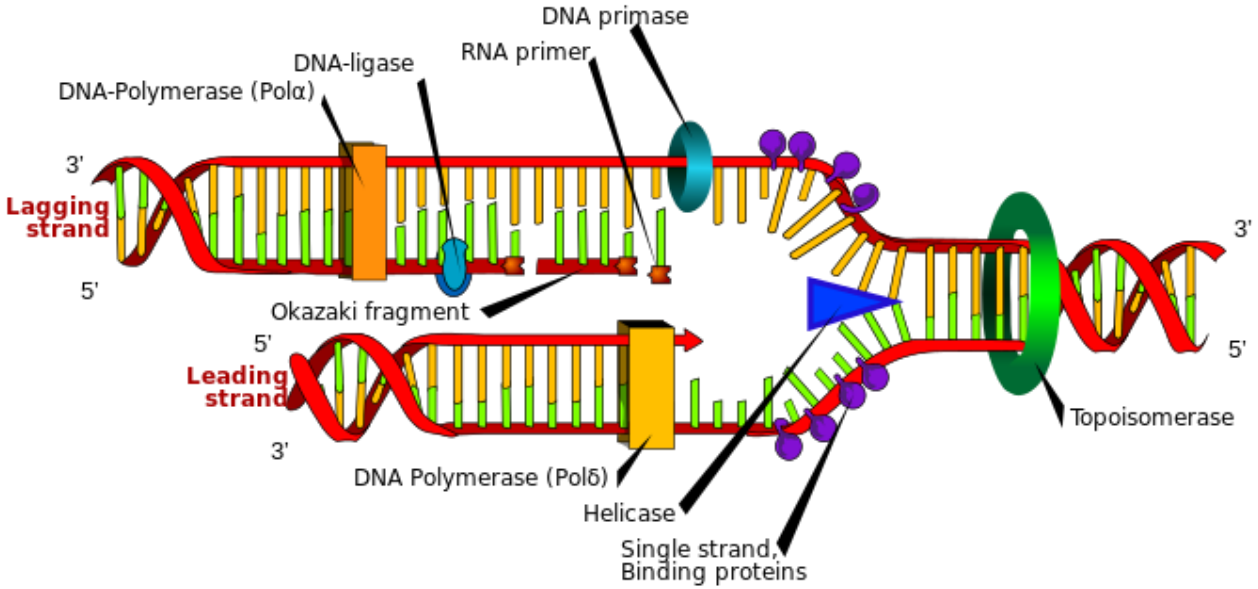
Cdc45 is an essential protein that together with Mcm2-7 and GINS forms the eukaryotic replicative helicase CMG. Cdc45 seems to be rate limiting for the initial unwinding or firing of replication origins. In line with this view, Cdc45-overexpressing cells fired at least twice as many origins as control cells. However, these cells displayed an about 2-fold diminished fork elongation rate, a pronounced asymmetry of replication fork extension, and an early S phase arrest. This was accompanied by H2AX-phosphorylation and subsequent apoptosis. Unexpectedly, we did not observe increased ATR/Chk1 signaling but rather a mild ATM/Chk2 response. In addition, we detected accumulation of long stretches of single-stranded DNA, a hallmark of replication catastrophe. We conclude that increased origin firing by upregulated Cdc45 caused exhaustion of the single-strand binding protein RPA, which in consequence diminished the ATR/Chk1 response; the subsequently occurring fork breaks led to an ATM/Chk2 mediated phosphorylation of H2AX and eventually to apoptosis.

Cdc45

- encoded by *CDC45L* gene (humans)
- rate limiting for initial firing of replication origins
- **Concentration of Cdc45** is carefully controlled:
  - 30.000-40.000 molecules per cell,
  - compared to 250.000 replication origins per cell
  - (tumor cells can show higher expression of Cdc45)
- **Cdc45 overexpression**:
  - leads to over-firing of replication origins (2x vs. control)
  - but
    - reduces fork elongation rate (~ 2x),
    - causes asymmetry of replication velocity
    - causes early S-phase arrest

DHX9

- helicase (Gene Cards ; Uniprot)
- unwinds double-stranded DNA and RNA 3' to 5'
- **Experiments**:
  - downregulated via RNAi (shRNA; "short hairpin")

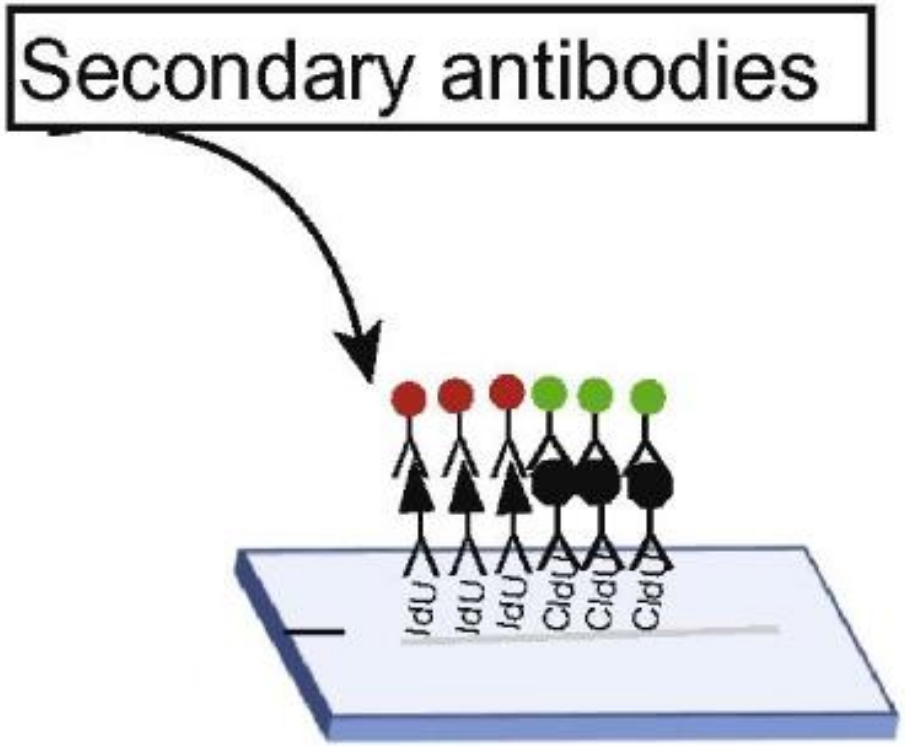


Visualization of replication

- DNA fibre fluorography
- visualization of progression of replication forks *in vivo*
- labelling of growing cells using 2 nucleotide analogues
  - **IdU** (5-Iodo-2'-deoxyuridine)
  - **CldU** (5-Chloro-2'-deoxyuridine)
  - applied one after another (**IdU** → **CldU**)
  - labelling pulse duration $\Delta t \sim 20$ min
- staining (antibodies)
- flourescence microscope
- image analysis

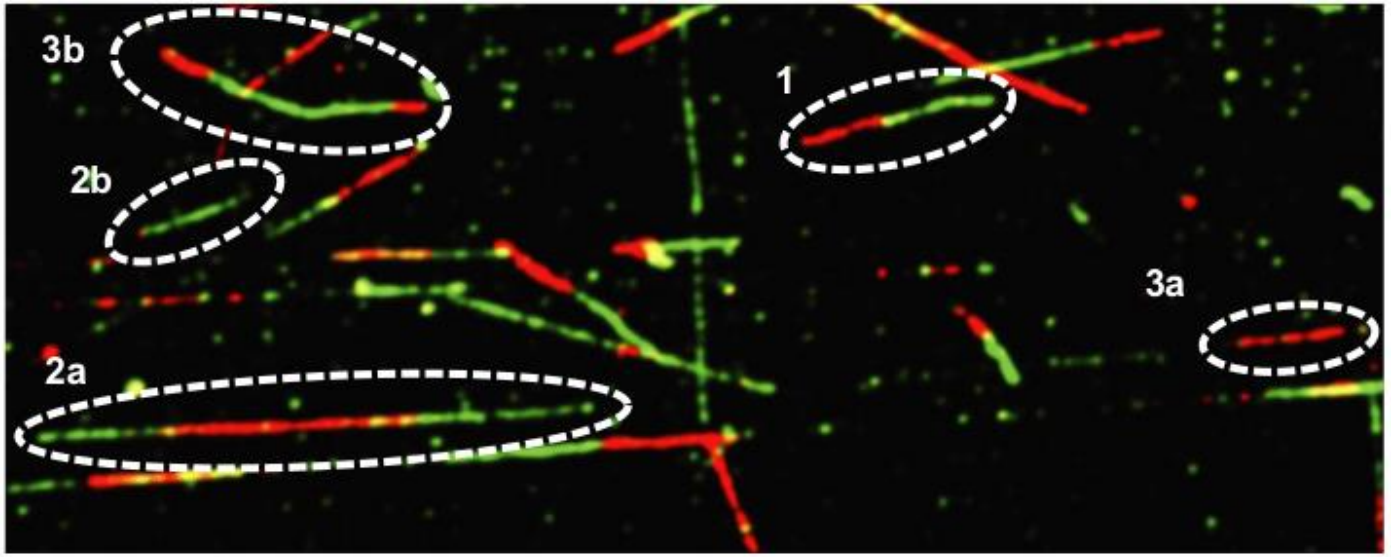
DNA fibre fluorography

Staining using antibodies



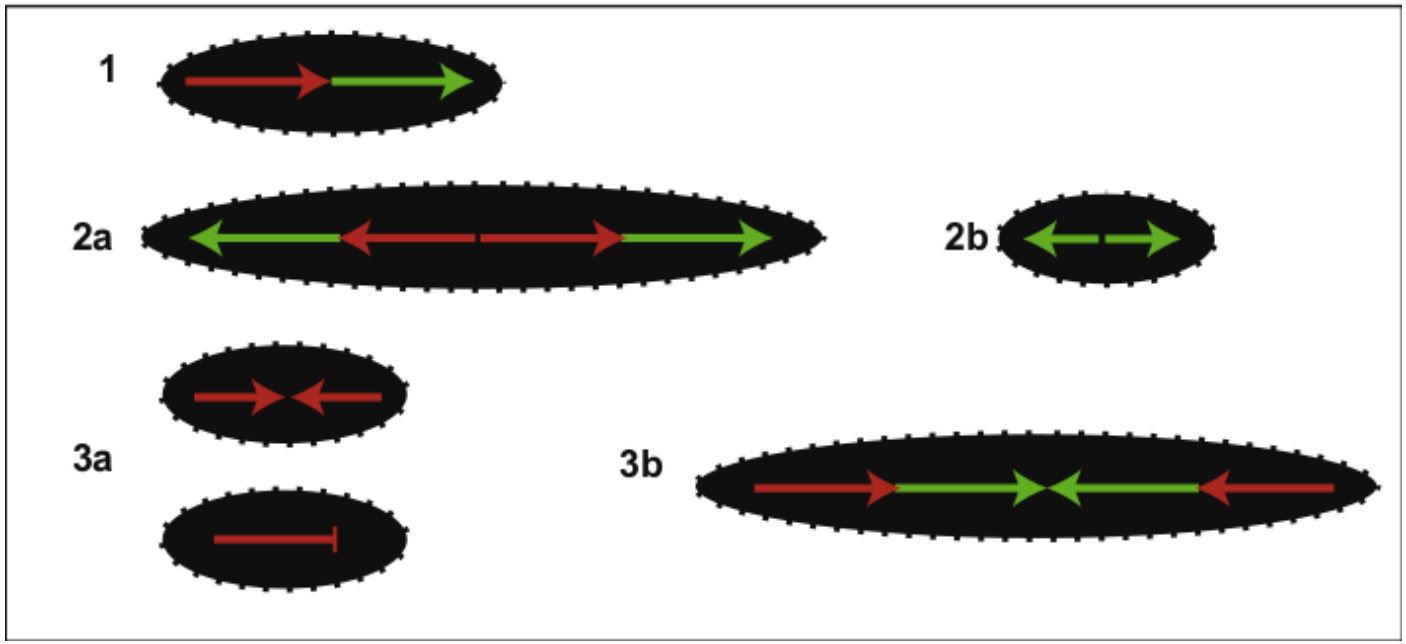
DNA fibre fluorography
Image analysis

- replication speed estimation by measuring fiber length
- replication velocity: 1 μm ~ 2.59 kb



- measuring fiber length is time consumimg!

DNA fibre fluorography



1: ongoing replication

2a: replication start during 1st labelling phase (**IdU**)

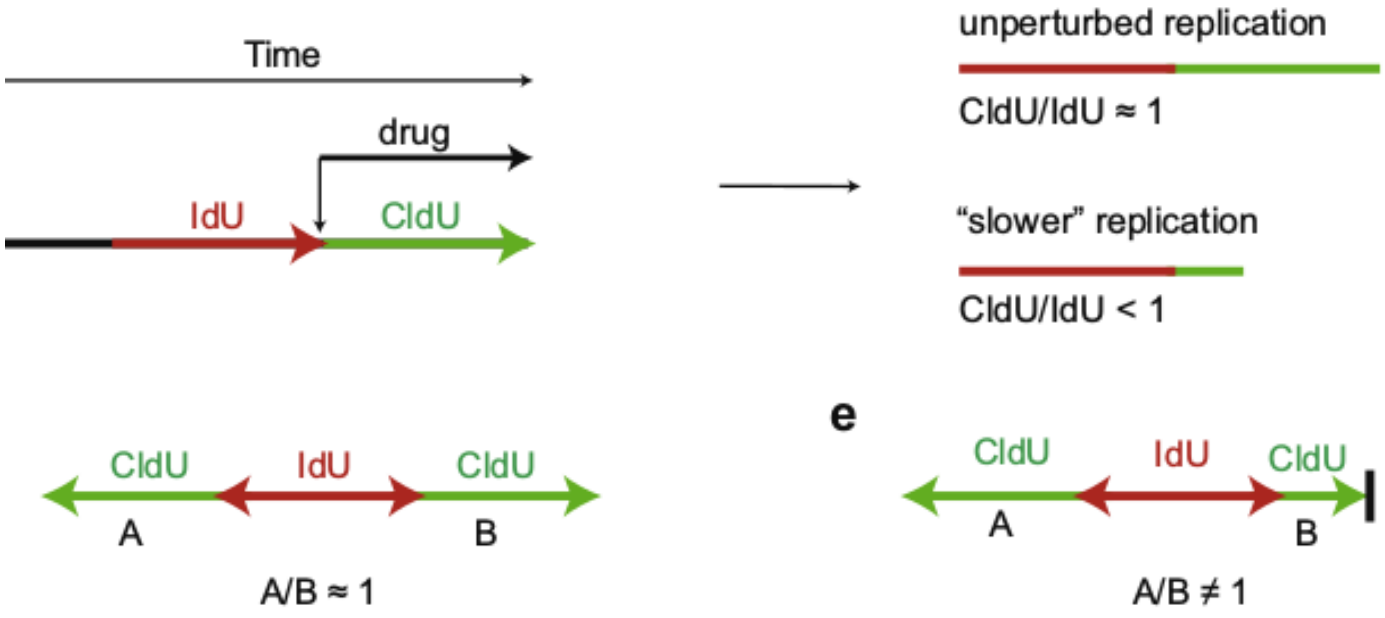
2b: replication start during 2nd labelling phase (**CldU**)

3a: stalled fork during 1st phase

3b: termination

Labelling variants

- known replicative inhibitors, stressors
- selective treatetment during one labelling pulse



fork symmetry

fork assymmetry (stalling)

Measurements

- of fork assymmetry (replication arrest, “stalling”)
- of replication rates (overall replication velocity)

	A	B
1	U2OS_LNA_elongation_20170826	
2	DHX9_LNA	SCR_LNA
3	0.6535844774	1.5159988612
4	0.5913870551	1.5663768084
5	1.2788713586	1.4836484814
6	0.8748565927	1.5159988612
7	0.4622669334	1.0191845324
8	0.19839887	2.4768477695
9	0.4249878065	0.7972752433
10	0.7534616175	0.4233832371
11	0.7702128101	1.5942756698
12	0.1519316065	0.7134154971
13	0.1947514824	0.692225565
14	0.5727633047	1.8436130651
15	0.9392796915	1.3823170899
16	0.1895775105	1.0263845241

significant difference?
(non-normally distributed)

Measurements

3 replicates

)HX9_LNA	)HX9_LNA	)HX9_LNA
0.6535844774	0.6535844774	0.6535844774
0.5913870551	0.5913870551	0.5913870551
1.2788713586	1.2788713586	1.2788713586
0.8748565927	0.8748565927	0.8748565927
0.4622669334	0.4622669334	0.4622669334
0.19839887	0.19839887	0.19839887
0.4249878065	0.4249878065	0.4249878065
0.7534616175	0.7534616175	0.7534616175
0.7702128101	0.7702128101	0.7702128101
0.1519316065	0.1519316065	0.1519316065
0.1947514824	0.1947514824	0.1947514824
0.5727633047	0.5727633047	0.5727633047

X_1 X_2 X_3

3 replicates

SCR_LNA	SCR_LNA	SCR_LNA
1.5159988612	1.5159988612	1.5159988612
1.5663768084	1.5663768084	1.5663768084
1.4836484814	1.4836484814	1.4836484814
1.5159988612	1.5159988612	1.5159988612
1.0191845324	1.0191845324	1.0191845324
2.4768477695	2.4768477695	2.4768477695
0.7972752433	0.7972752433	0.7972752433
0.4233832371	0.4233832371	0.4233832371
1.5942756698	1.5942756698	1.5942756698
0.7134154971	0.7134154971	0.7134154971
0.692225565	0.692225565	0.692225565
1.8436130651	1.8436130651	1.8436130651
1.3823170899	1.3823170899	1.3823170899
1.0263845241	1.0263845241	1.0263845241

Y_1 Y_2 Y_3

t-Test [CLT]

Bootstrap approach

Analyzing the difference of means based on one sample for each condition → more efficient

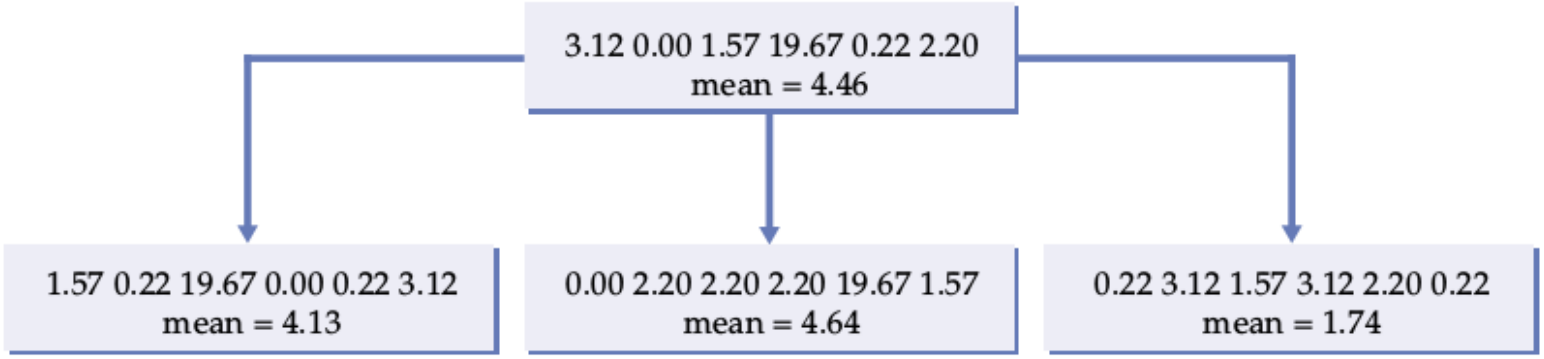
)HX9_LNA
0.6535844774
0.5913870551
1.2788713586
0.8748565927
0.4622669334
0.19839887
0.4249878065
0.7534616175
0.7702128101
0.1519316065
0.1947514824
0.5727633047

)HX9_LNA
0.6535844774
0.5913870551
1.2788713586
0.8748565927
0.4622669334
0.19839887
0.4249878065
0.7534616175
0.7702128101
0.1519316065
0.1947514824
0.5727633047

Bootstrap analysis
of the difference of means

Bootstrap idea

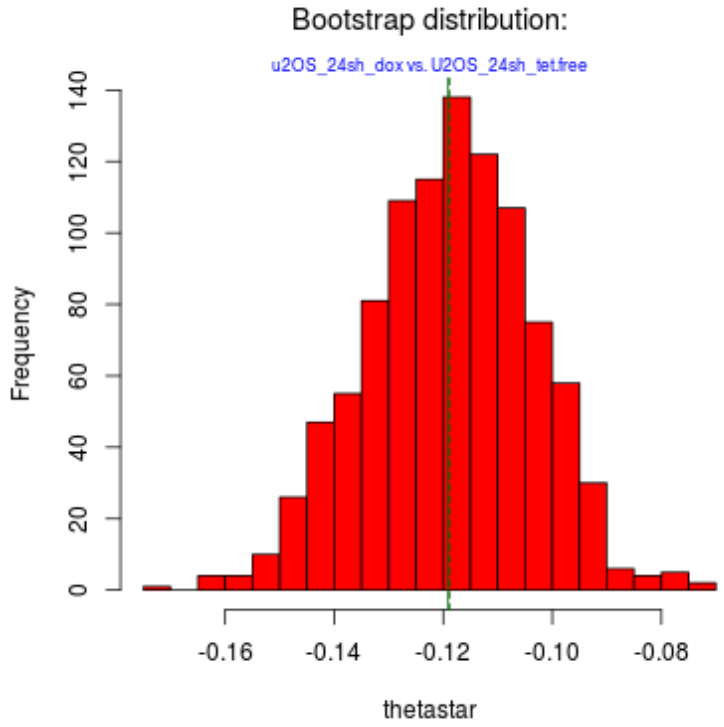
- mimic the sampling distribution of a statistic (mean)
  - get info about spread of the statistic
  - start with one original sample
  - resample from this sample with replacement, ~ 1000 times
  - calculate the statistic for each resample
- → bootstrap distribution
  - agrees with sampling distribution in shape and spread

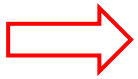


Graphic taken from: Moore, McCabe, Craig;
Introduction to the Practice of Statistics

Bootstrap distribution

```
Repeat 1000 times {
  Draw a resample with replacement from the data.
  Calculate the resample mean.
  Save the resample mean into a variable.
}
Make a histogram and normal quantile plot of the 1000 means.
Calculate the standard deviation of the 1000 means.
```





Confidence interval

Bootstrap t-confidence interval

$$I_\theta = \bar{x} \pm t_{\alpha/2} \cdot SE_{boot}$$

x: statistic of the original sample

$t_{\alpha/2}$: quantile for t-distribution, confidence level α

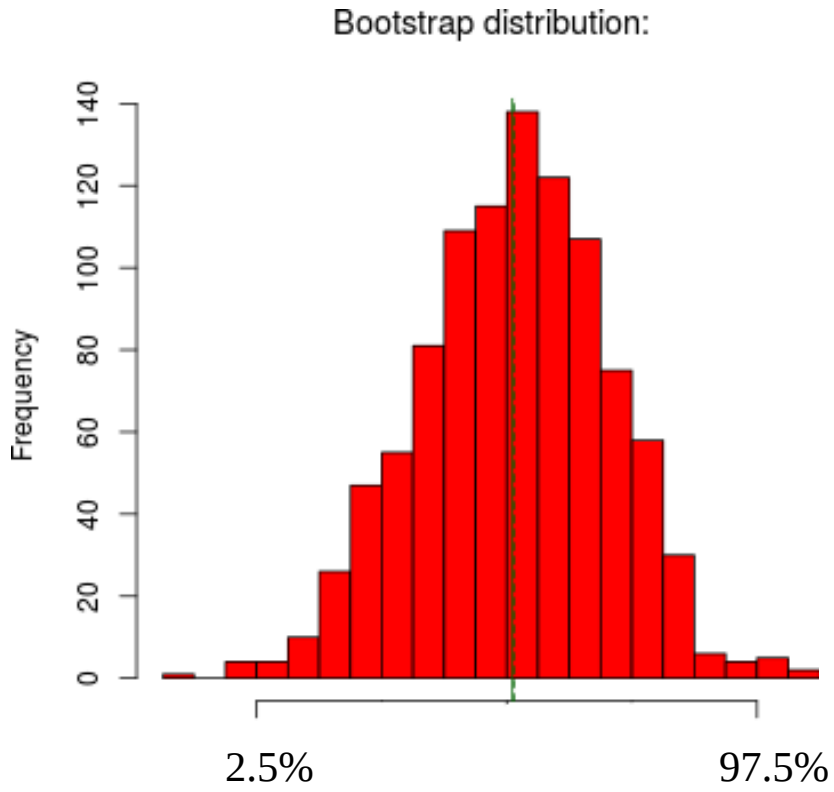
SE_{boot}: bootstrap standard error

$$SE_{boot} = \sqrt{\frac{1}{B-1} \cdot \sum_{i=1}^{B} (\theta_i - \bar{\theta})^2}$$

B: number of resamples

- works well if the bootstrap distribution is approximately normal and if the bias is small
- SE_{boot} is directly taken from the bootstrap values (θ_i)

Bootstrap percentile confidence interval



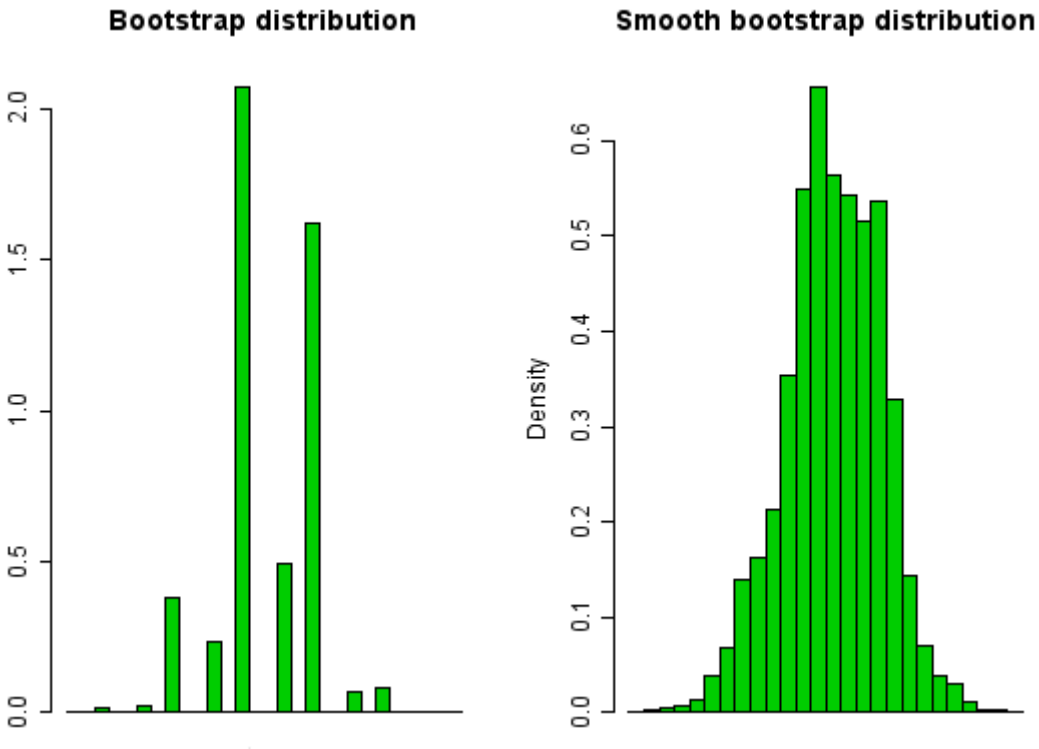
- mark off the central 95% → 95% confidence interval
- does not ignore skewness, but bias must be small
- R: I = quantile(thetastar, probs = c(0.025, 0.975))

BCa confidence interval

- More accurate, recommended!
- **B**ias-**C**orrected **a**ccelerated CI
- modification of percentile CI
- adjusts for bias and skewness
- accurate unless the sample size is very small

Still, small samples do not work!

Bootstrap inference from small samples is unreliable !



Bootstrap in R

- **loop!**
  - sample(x, length(x), replace = TRUE)
- **library(boot)**
  - meanDiff = function(df, iv) {…}
  - b.out = boot(data = df, statistic = meanDiff, R = 999)
  - boot.confint = boot.ci(b.out, type = "bca")
- **library(bootstrap)**
  - theta = function(x) { … }
  - res = bcanon(x, 999, theta)
- **library(wBoot)**
  - boot.two.bca(x, y, mean, null.hyp = 0, R = 999, conf.level = 0.95)

Software

- R-Shiny-app (not 100% finished ...)
- Bootstrap + a little bit more
- from R console:
  - > library(shiny)
  - > runApp("/your_path/shiny_multtest")
- from Linux command line:
  - > R -e "shiny::runApp('/your_path/shiny_multtest')"
  - Listening on http://127.0.0.1:7789 → open in browser
- from Windows:
  - via R console: as above
  - via RStudio: open the app.R in /your_path/shiny_multtest/ and click the "Run App" button
- terminates with Ctrl-C

T-test

- **Assumptions**:
  - both samples normally distributed
  - independent sample points

- from R console:
  - > library(shiny)
  - > runApp("/your_path/t_test_tabbed")
- from Linux command line:
  - > R -e "shiny::runApp('~/your_path/ t_test_tabbed')"
  - Listening on http://127.0.0.1:7789 → open in browser
- from Windows:
  - via R console: as above
  - via RStudio: open the app.R in ../ t_test_tabbed/ and click the "Run App" button
- terminates with Ctrl-C