

General influence of mRNA secondary structure on splicing

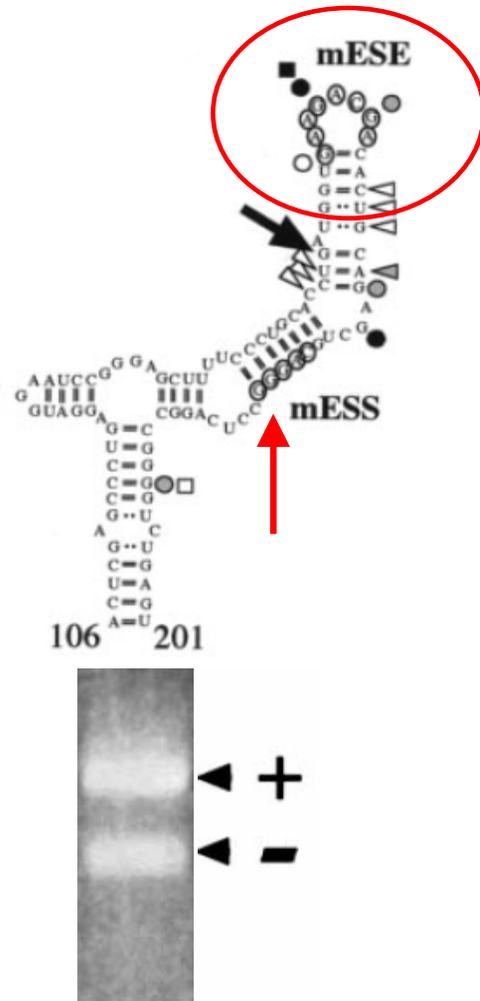
(and what NAGNAGs have in common
with sex, drugs and death)

Michael Hiller
Bioinformatics Group
University of Freiburg, Germany

Secondary structure of mRNA

- mRNA is often considered as a linear sequence, but
- structural elements: IRE, IRES, SECIS, A to I editing
 - influence of secondary structures on splicing:
 - exon 10 skipping of *MAPT* is influenced by a stem structure that sequesters the donor
 - regulation of mutually exclusive exons in *FGFR2*
 - cluster of 48 mutually exclusive exons in *Drosophila DSCAM*
 - SR proteins / hnRNPs have “single-stranded RNA binding domains”
 - bind hairpin loops, e.g. Nova1, hnRNP A1, SRp55

Fibronectin EDA exon



General trend for splicing motifs to be single-stranded?

1. Data set

Experimentally verified splicing motifs

- AEDB motif database:
 - splicing motifs in their natural sequence context
 - no predicted motifs, splicing-relevant SNPs
 - only motifs shorter than 10 nt
- final set of 77 motifs

2. How to measure single-strandedness?

1. probability that an mRNA part is completely unpaired

$$PU = e^{-\frac{E_{all} - E_{unpaired}}{kT}}$$

2. expected fraction of unpaired bases in an mRNA part

$$EF = 1 - \frac{\sum_{i=a}^b \sum_{j=1}^N P_{i,j}}{b-a+1}$$

- the higher PU or EF, the higher the single-strandedness
- allow comparison for motifs of the same length

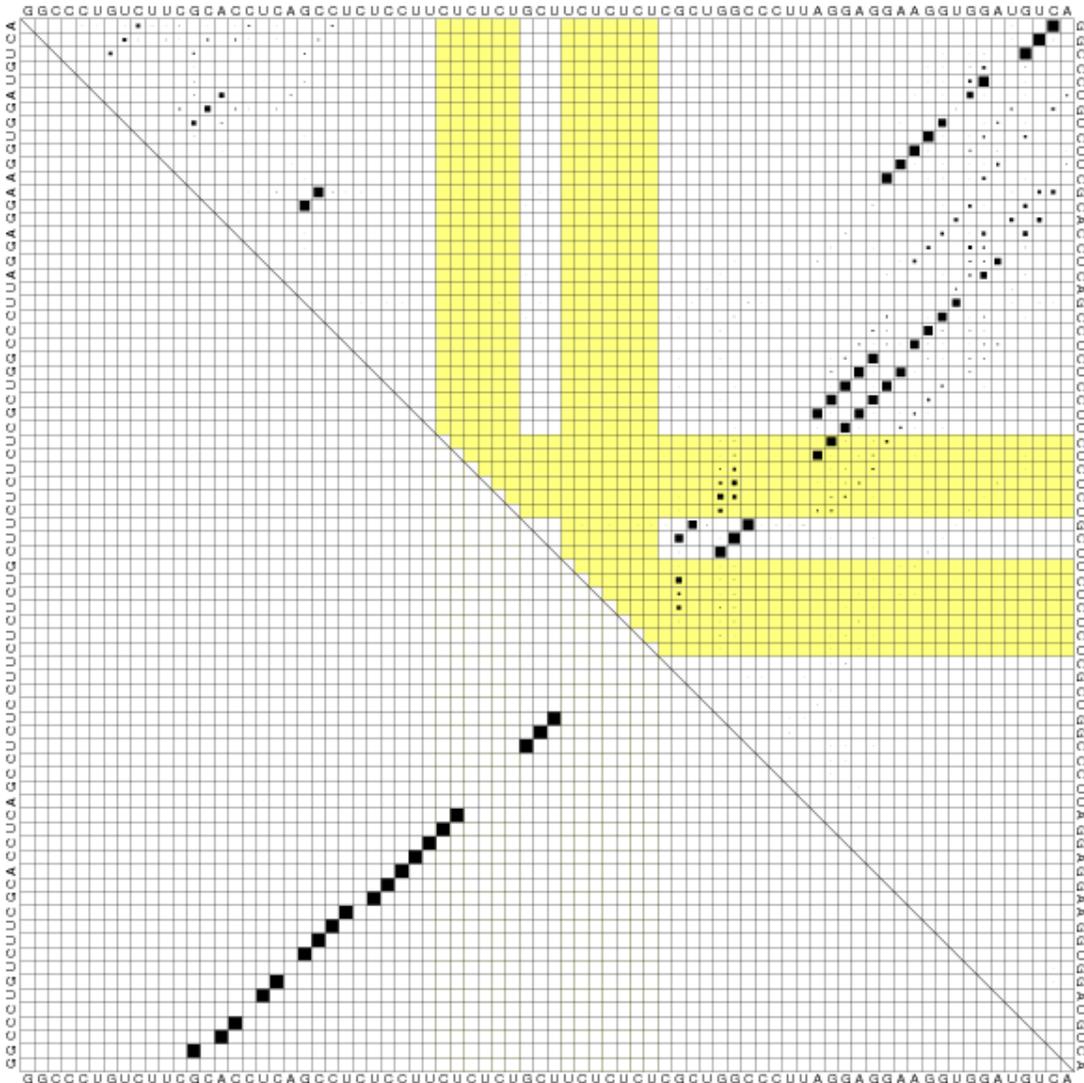
Example: PTB binding sites

GGCCCUUGUCUUCGCACCCUCAGCCUCUCCUUUCUCUCUGCUUCUCUCUCGCGCUGGCCCUUAGGAGGAAGGUGGAUGUCA
 (((...(((.....))..(((.....((.....)))..)))..))).. -18.40
 (((...(((.....))..(((.....((.....)))..)))..))).. -18.16
 (((...(((.....))..(((.....((.....)))..)))..))).. -18.10
 (((...(((.....))..(((.....((.....)))..)))..))).. -18.10
 (((...(((.....))..(((.....((.....)))..)))..))).. -18.10
 (((...(((.....))..(((.....((.....)))..)))..))).. -17.86

CTCTCT:

EF = 0.93

PU = 0.71



TCTCTCT:

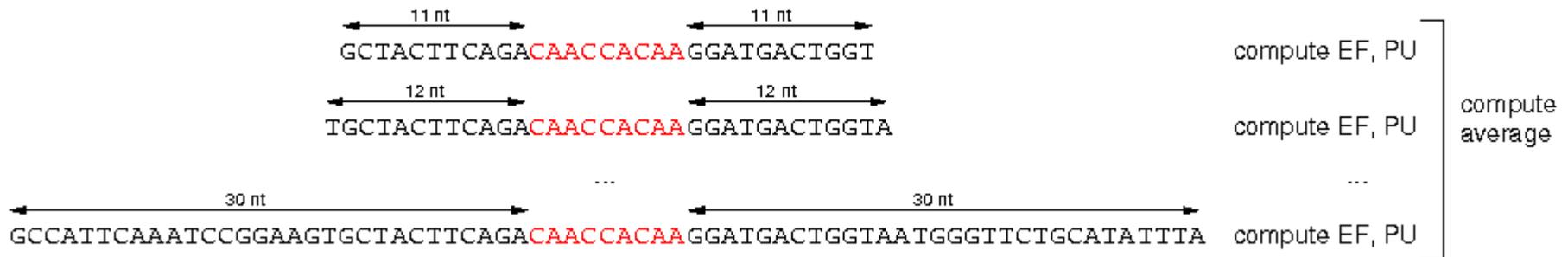
EF = 0.95

PU = 0.60

3. In which region is mRNA free to fold?

- folding is affected by
 - protein binding
 - co-transcriptional structure formation
- experimental evidence that folding is limited to ≈ 50 nt
 - short range base pairs are more likely
- consider symmetrical context lengths 11 – 30 nt
- compute average EF and PU

GCCATTCAAATCCGGAAGTGCTACTTCAGACAACCAAAAGGATGACTGGTAAATGGGTTCTGCATATTTA



Results and Statistics

- 77 experimentally verified motifs:
 - average EF: 0.65
 - average PU: 0.25
- null model 1: pick random motif from flanks

AACTTCCTTTATTTTCCTTACAGGGTTT**TAGACA**AAATCAAAAAGAAGGAAGGTGCTCAC



AACTTCCTTTATTTTCCTTACAGGGTTTTAGACAAAATCAAAA**GAAGGA**AGGTGCTCAC

- generate 100 random sets
 - compute average EF and PU value for each set
 - P-value = [number of random sets with higher EF or PU] / [100]
- average EF: 0.59 P<0.01
- average PU: 0.15 P<0.01

Statistics

- null model 2: copy motif to the flanks

AACTTCCTTTATTTTCCTTACAGGGTTT**TAGACA**AAATCAAAAAGAAGGAAGGTGCTCAC



AACTTCCTTTATTTTCCTTACAGGGTTTTAGACAAAATCAAAA**TAGACA**AAGGTGCTCAC

→ average EF: 0.60 P<0.01

→ average PU: 0.18 P<0.01

- null model 3: dinucleotide shuffling

AACTTCCTTTATTTTCCTTACAGGGTTT**TAGACA**AAATCAAAAAGAAGGAAGGTGCTCAC



TTCTAATTTATCCTAACATCAGCTAAT**TAGACA**TTCTTCCTTGCTAAAACAAATTACTA

→ average EF: 0.59 P<0.01

→ average PU: 0.15 P<0.01

Statistics

- null model 4:
 - take 10000 random motifs from exons
 - compare with verified exonic motifs

→ average EF: 0.57 t-test: $P < 0.057$

→ average PU: 0.12 t-test: $P < 0.018$
- null model 5:
 - take 10000 random motifs from introns
 - compare with verified intronic motifs

→ average EF: 0.58 t-test: $P < 0.001$

→ average PU: 0.15 t-test: $P < 0.001$

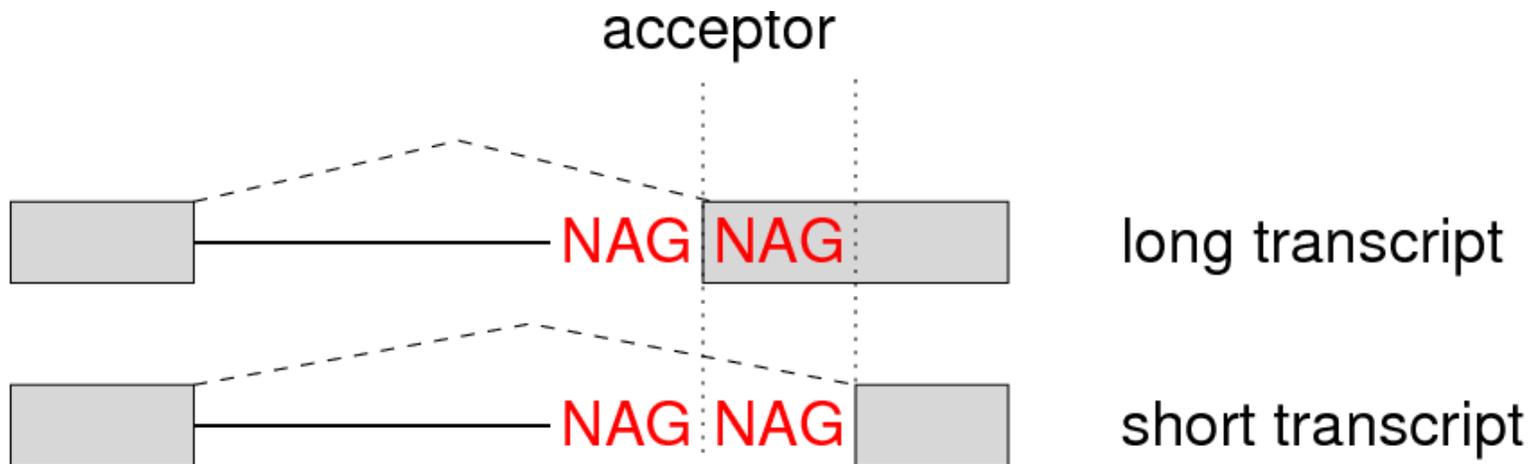
Conclusion

- verified splicing motifs are more single-stranded
- attributed to the flanks
- selection pressure on a coding exon:
 - preserve coding sequence
 - preserve splicing signals
 - preserve an appropriate structural context for splicing motifs
- another piece for the ‘mRNA splicing code’

Part 2

what NAGNAGs have in common with

- sex
- drugs
- and death (not rock'n roll)

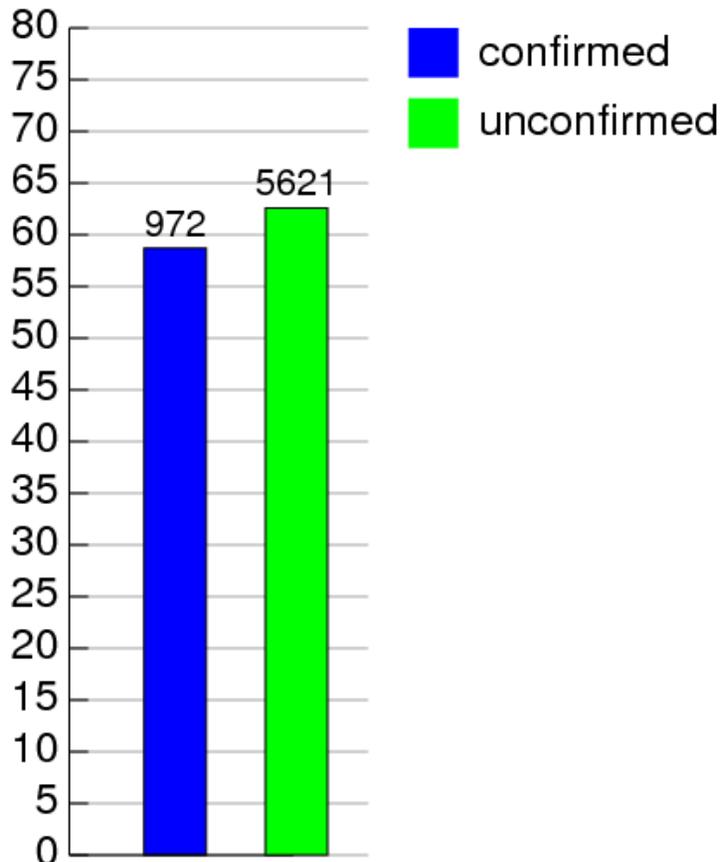


N = A,C,G,T

NAGNAG conservation

- compare alternatively spliced NAGNAGs (confirmed) with not alternatively spliced NAGNAGs (unconfirmed)

percent conserved in mouse

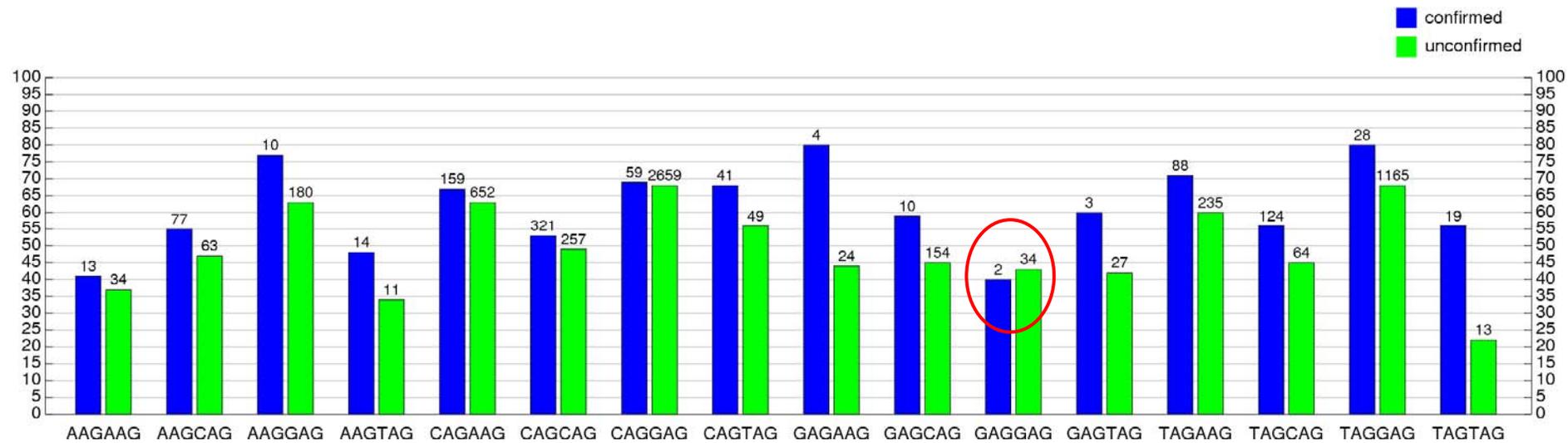


Fisher's exact test:

$P < 0.0024$

→ significant difference

NAGNAG Conservation



- for 15 of the 16 NAGNAG motifs confirmed are more conserved than unconfirmed
- Cochran-Mantel-Haensel test that corrects for the influence of the motif

odds ratio: 1.35 with $P < 0.000007$

→ confirmed NAGNAGs are more conserved than unconfirmed

How can that be?



Simpson's paradox (Simpson 1951, J Royal Stat Soc):

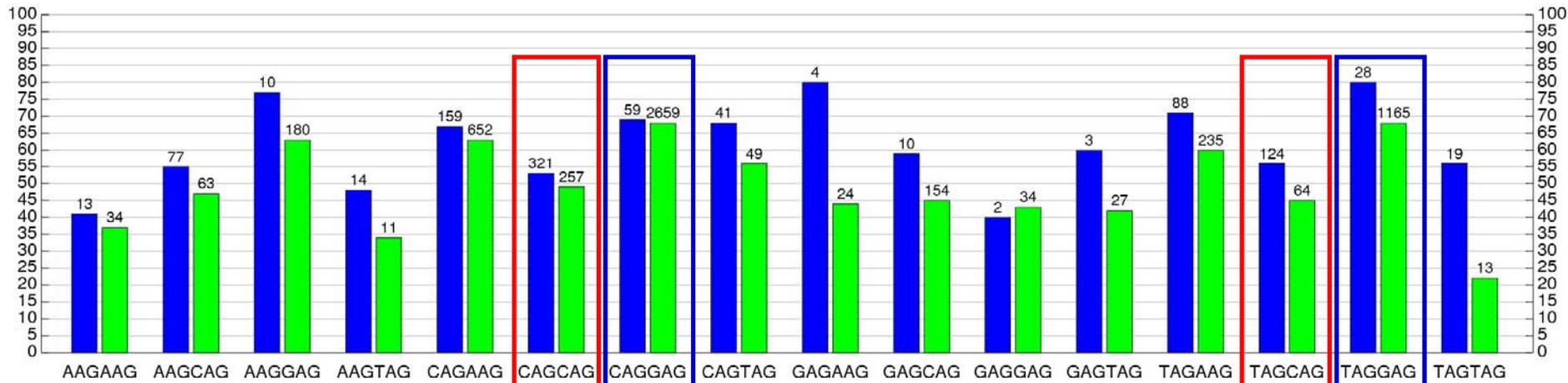
a trend observed between two groups seem to be reversed when the groups are splitted into subgroups

occurs since

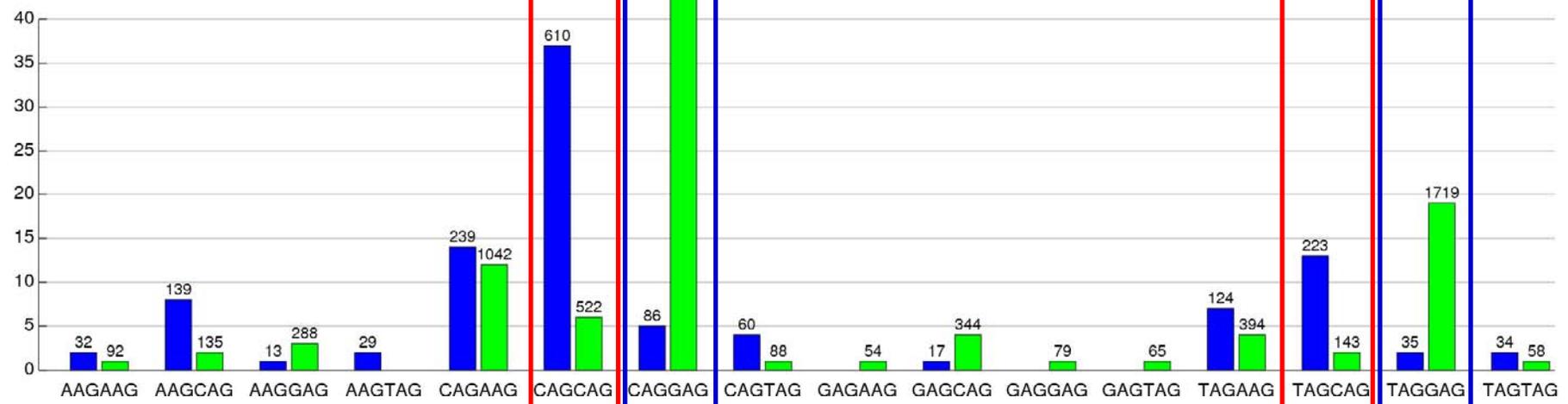
1. the conservation level varies considerably between the motifs
2. motif distribution for confirmed and unconfirmed is different

Simpson's paradox

conservation



motif distribution



conservation level:

CAGCAG + TAGCAG ~ 50%

CAGGAG + TAGGAG ~ 70%

motif distribution:

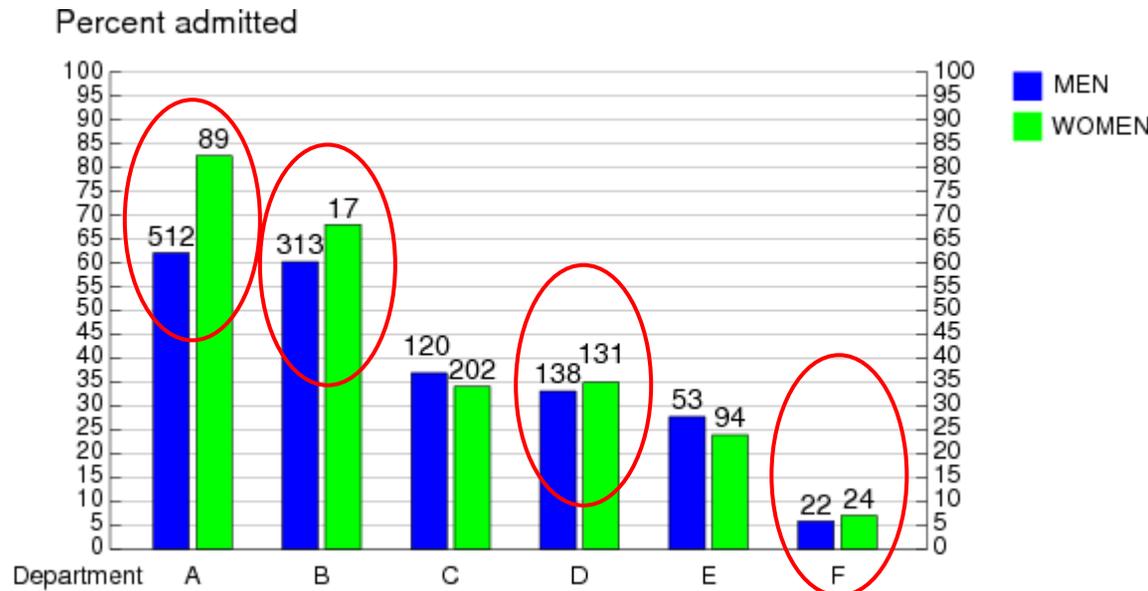
CAGCAG + TAGCAG make up 50% of confirmed

CAGGAG + TAGGAG make up 63% of unconfirmed

Berkeley sex bias case

- UC Berkeley in the early 70's was suspected to have a bias against women applications
 - 44% of 8442 men were admitted
 - 35% of 4321 women were admitted
- Fisher's exact test: P-value = 0.0000000000000000000000000000035

→ 4 of 6 largest departments had a bias **against men**

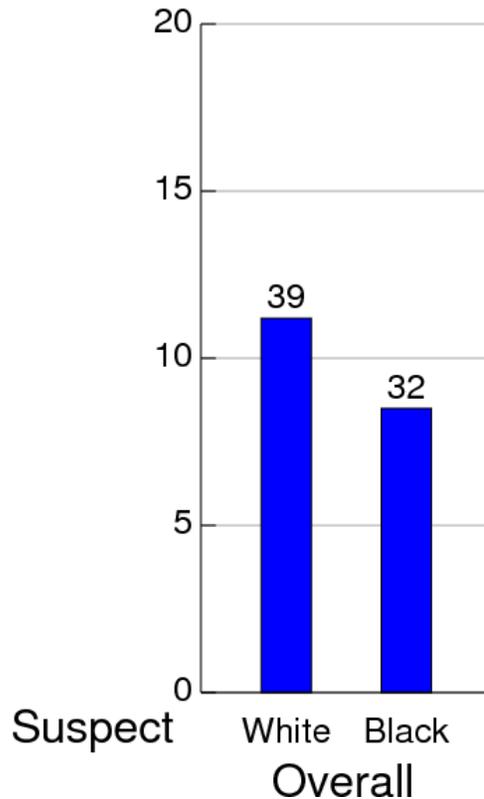


- SOLUTION:
- departments have **different** admission rates
- women tended to apply to departments with **low** rates of admission
- men tended to apply to departments with **high** rates of admission

Death sentence data

Is the application of the death sentence racially motivated?

Percent with death sentence (data from Florida)

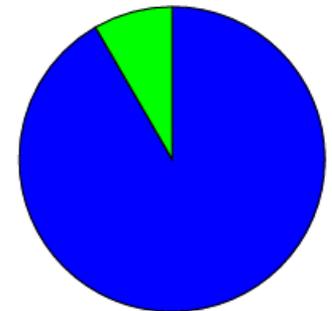


↓
NO

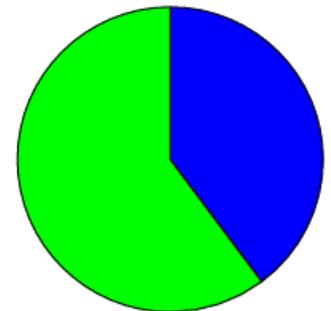
↓
ES

■ white victim
■ black victim

white suspect



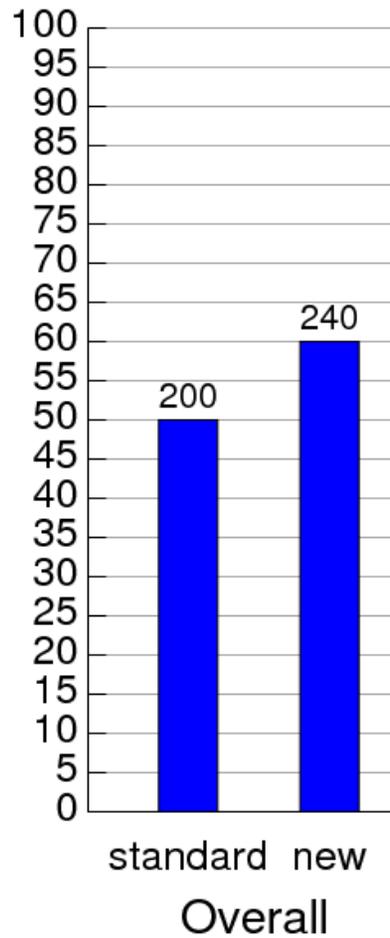
black suspect



Drug data

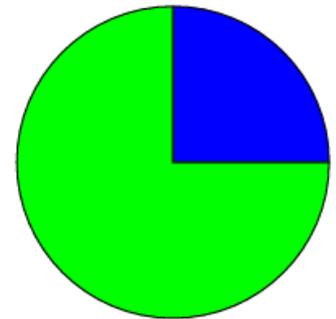
Is a new drug better than the standard drug?

percent healed

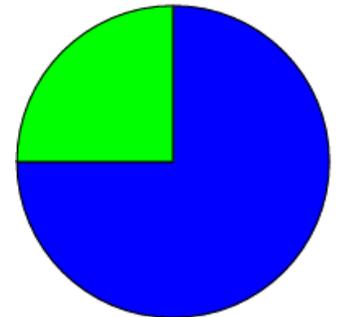


■ standard
■ new

women



men



Conclusion

What can we learn?

Whenever the observed trend goes into the wrong direction
→ try to find a confounding variable that reverses this trend !

	variable
NAGNAG	motif
Berkeley sex bias	department
drug data	sex
death sentence	race of victim

[unclear whether from Churchill or Goebbels]

Der Vortrag gibt dem Forscher Kraft,
verschafft er ihm ein wenig Schlaf.