

Биоинформатика в 
что ДНК говорит нам о
жизни и о ее истории

М. Гельфанд

50-лет ИППИ

29.12.1961 - 7.12.2011

2001: Лаборатория № 6 (В.А.Любецкий)

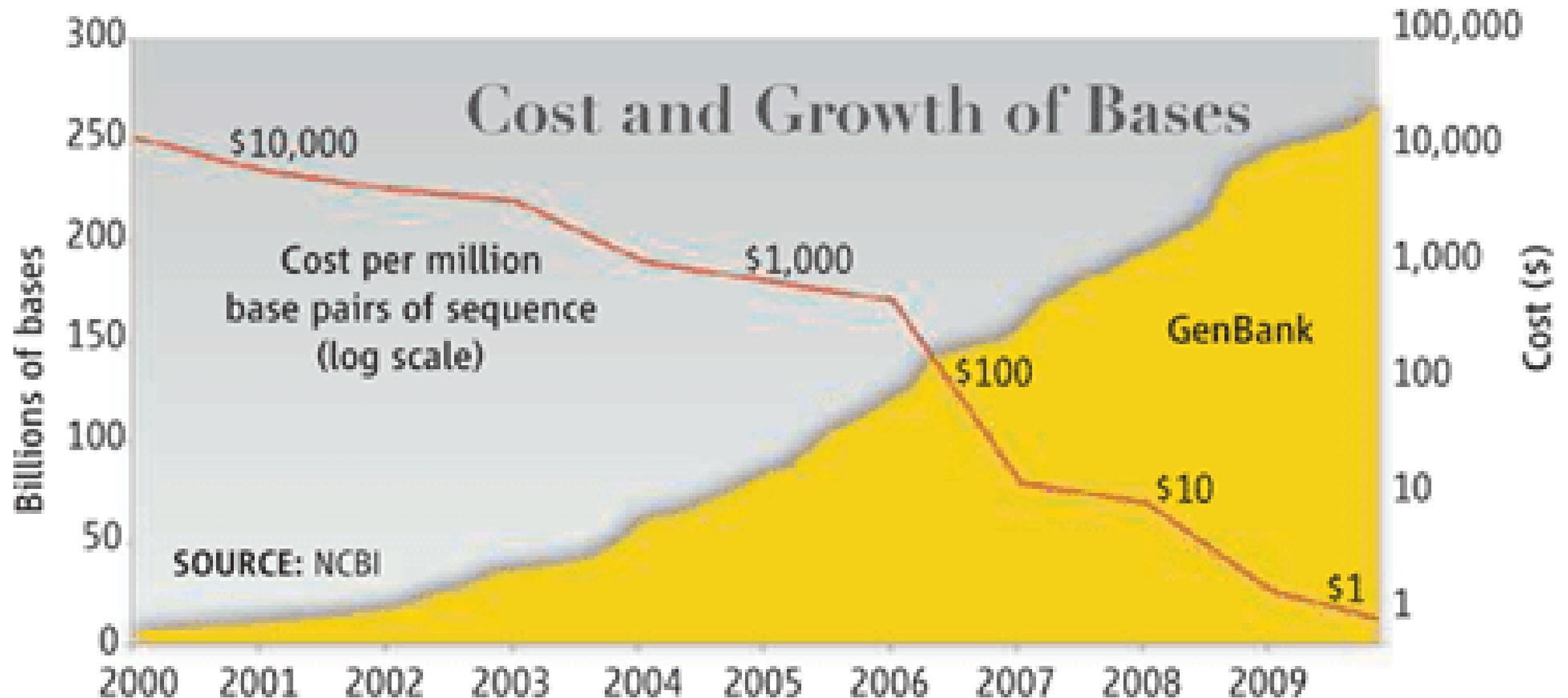
2003



2008: Сектор № 4 (Г.А.Базыкин)

2010: Сектор № 6 (Д.А.Родионов)

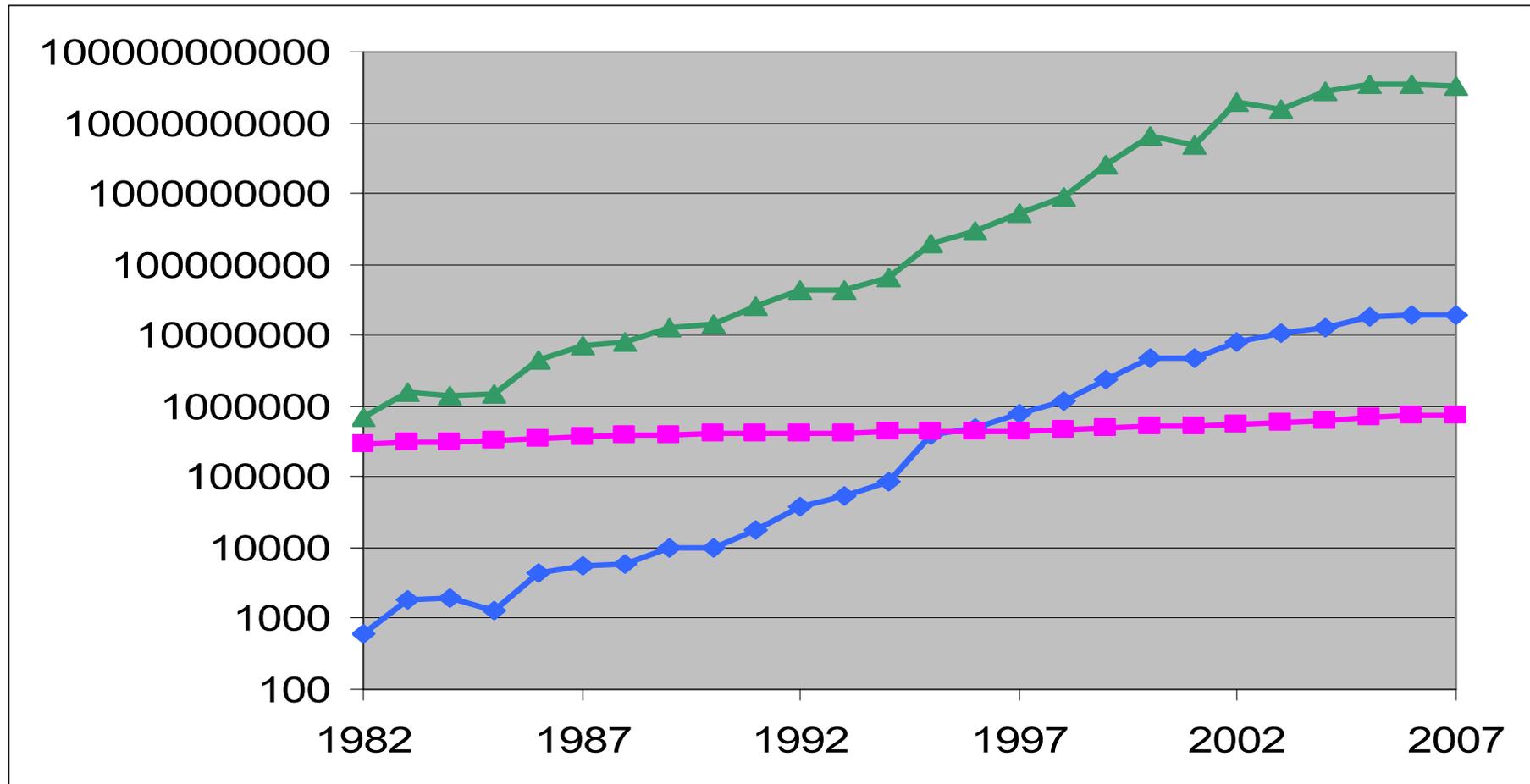
So much for so little. The decline in sequencing costs (red line) has led to a surge in stored DNA data.



Will Computers Crash Genomics?

Science 11 February 2011:
vol. 331 no. 6018 666-668

Рост объема нуклеотидных последовательностей



красный - статьи (PubMed)
голубой - фрагменты ДНК (GenBank)
зеленый - нуклеотиды (GenBank)

из 18 млн. статей в PubMed, ~675 тыс. имеют ключевое слово "bioinformat* OR comput*"

Полные бактериальные геномы

30

количество геномов, полная последовательность которых была опубликована в течение года

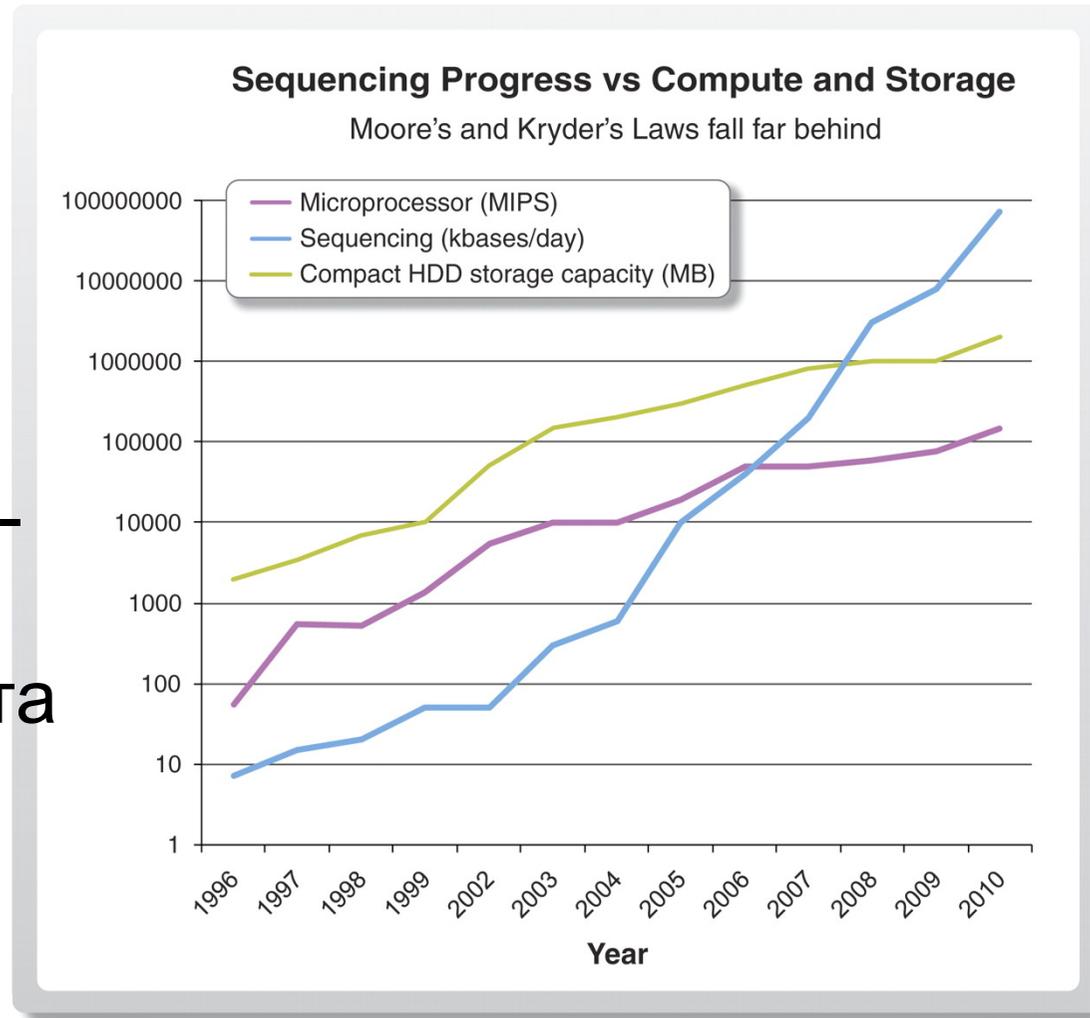
85

общее количество геномов с известной полной последовательностью



Fig. 1 A doubling of sequencing output every 9 months has outpaced and overtaken performance improvements within the disk storage and high-performance computation fields.

на самом деле, все-таки экспонента

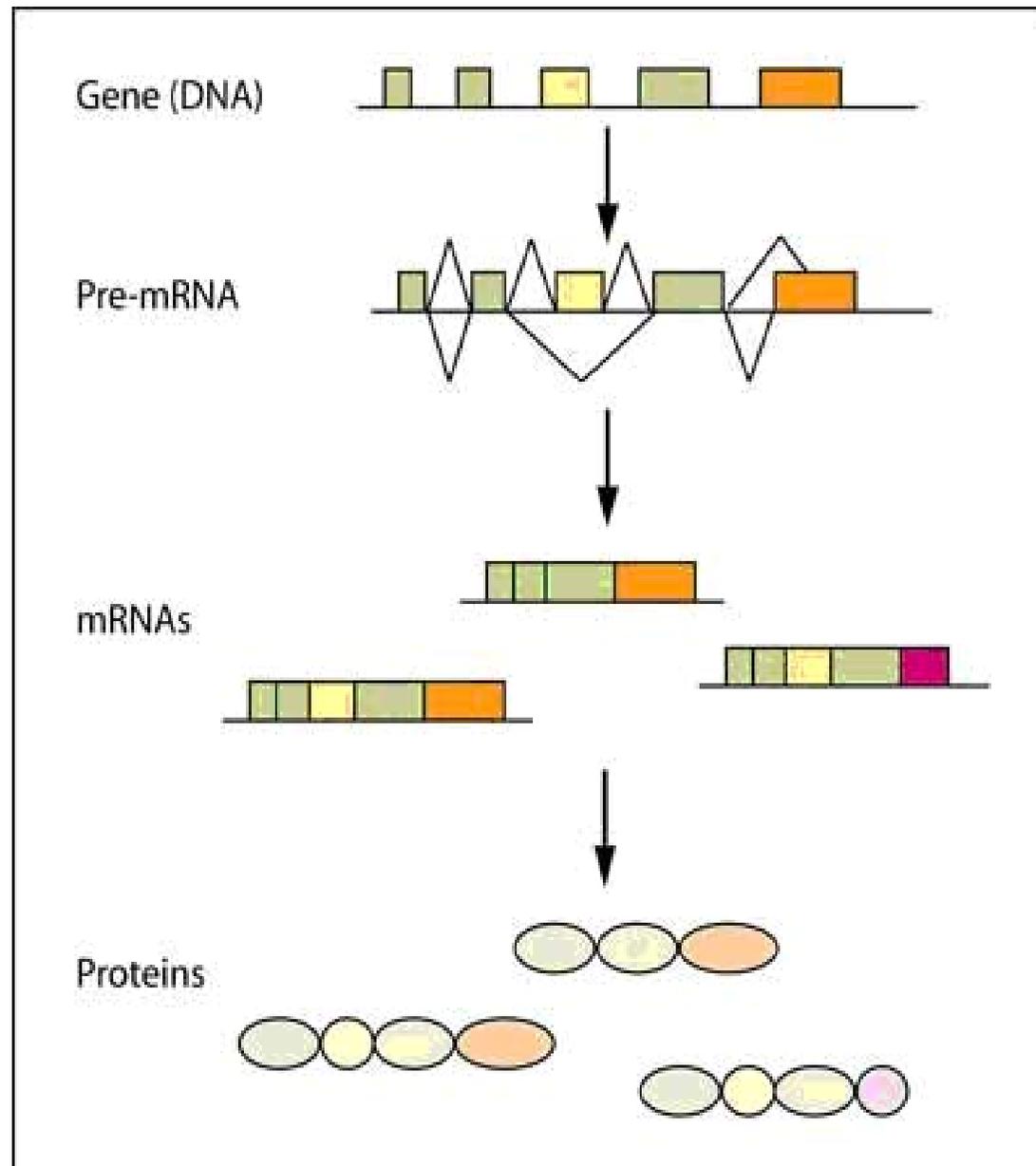


S D Kahn Science 2011;331:728-729



0.0001% генома человека

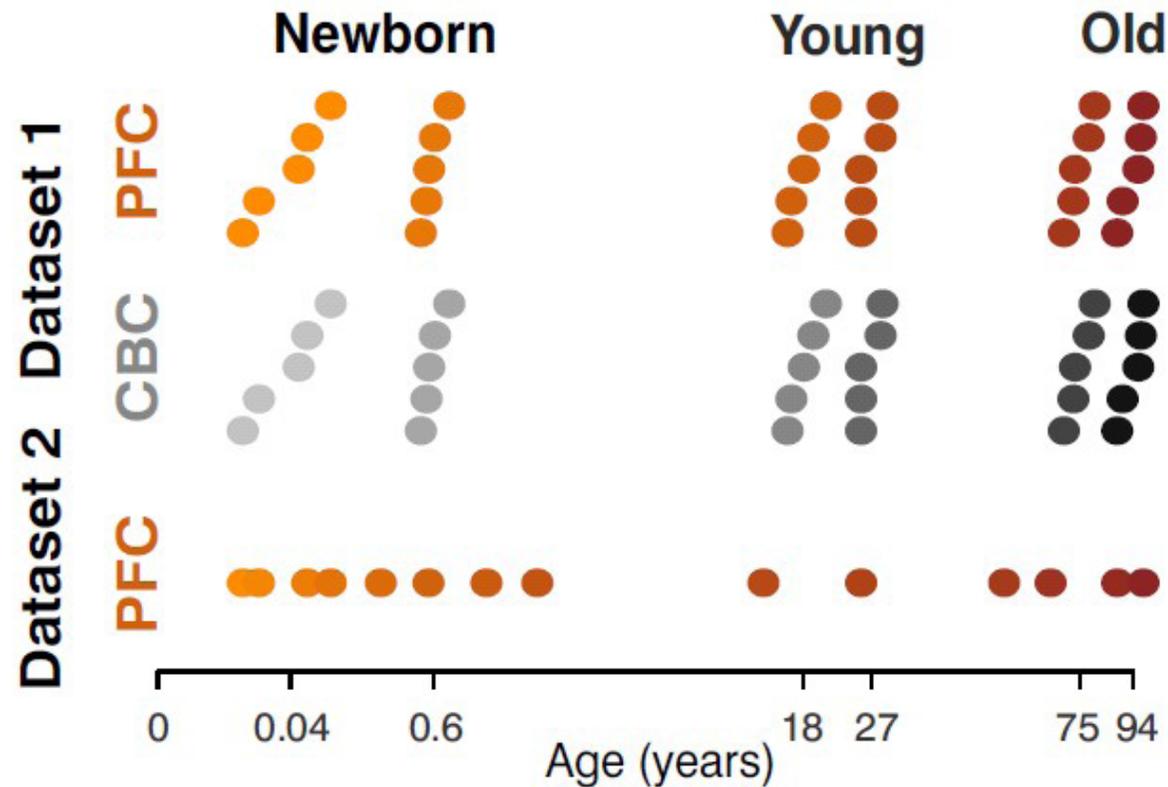
cgtgcac ttctgaaggacttcaggtac cggcgtgccc cgcgtcctac tgtcc gcctgctcgc gtcctgggtgccgcc tctgagtagggcgggagagg
CAGCCAA GCGGGAGCTGATGGCTGCGC CGAGG GCGGG GCGGG GTGCA GGCTGGAGCC TTCGG GCATG GCGGG CTTTGGGGG CATT CGCTG GGGGAGGAG
ACCCCGTTTGAC CCCTGACCTC CGGGC CCTGCTGACGTCAGGA AACTTCTGAC CCCC GGGCC GAGTGACTTATGGGACCCCCAGTCTCTGGGCCCGTTG
TCTGTTGGGTC ACTGAACCCC GAGCATGCCTGACGTCTGGGACCCC GGGTC CCCGGGCACA ACTGACTGCGGTGAC CCCAGATAC CAGGACCCGGGAGG
CCTCAGAGAACTCTGGAACCCGTTCCG GCGCGTGGCTGGCGGTGGCGCTGGGCGCTGGGGGGCAGTGCTGTGTTGTTGTGGGGC GGGGGTCGGGGTCC
TCCGGCC GTCCTCGCCG CCGTC CCTAGCCC GC CCCC GCTTC TCCCC GGAGTCAGTACA AACTTCATC GCAGATGTGGTGGAGAAGACAGCA CCTGC CGTG
GTCTATATCGAGATCCTGGACC Ggtaa tgggtggggtagacc gggagggactgaagc cacagcctggaggggc gggcgggtaggagggtcagagcc tcc
cttatctgtgctttccc tccatttcagGCACC CTTTCTGGG CCGCGAGGTC CCTATCTCGAACGGC TCAGGATTCGTGGTGGCTGCCGATGGGCTCATT
GTCACCAACGCC CATGTGGTGGCTGATCGGCGCAGAGTCCGTGTGAGACTGCTAAGC GCGGACACGTATGAGGCCGTGGTCA CAGCTGTGGATCCC GTGG
CAGACATCGCAACGCTGAGGATTCAGACTAAG gtgggggctggggtagggcaggctctgggtgagctgcttattgctcgcac tcttcagatgacaggtct
cttttac ccatttccc ttaggagcctctccc cacgc tgcctctgggacgctcagctgatgtccggc aaggggagtttgtgtgtgc catgggaagtccct
ttgcaactgcagaacacgatcac atccggcattgttagctctgtctcagcgtcc agccagagac ctgggactcc cccaa accaa tgtggaatacattc aac
tgatgcagctattgatgtgctcctgtaggagagaaatgac aatgatggggggagggggagggc tgtgtggtac aagcaaccaactgatatatggtg
atgagcc tatatagagc ttaggctgcaaaaatgtggc cacttattca tgggc tgaga aagaagagaa tttggagaaagtacc taca tccgtggtatgcccc
cagacttagaatccccagatctctttc atgttttctc cttgtcctac agTTTGGAAA CTCTG GAGGT CCCCTGGTTAACCTG gtgagtgagacatcctc
cttccaagaatc cctgc cccaggtcagtggtgggaagggtaggtttcc cctaa ttcaaggatgtttggtcaagtttctgagcagttc tttgtggctatct
ctcaata tccaaccagatctcc ccaac acttgctggtacttt tgttc gggtgcccc atccc ctactatttgtttaggctagggaactgggggctgtatc
cctgcagGATGGGGAGGTGATTGGAGTGAACA CCATGAAGGT CACAGCTGGAATCTC CTTTGCCATC CCTTC TGATC GTCTTCGAGAGTTTCTGCATCGT
GGGGAAAAGAAG Agtga gcctgcctta tggggaaacgggttc cttaatgtggtgga aataggggaaagggca ttcagtgggacttc ctggagggtggtct
actgggagaagaggggaggyaaggaagatgt agctgggtggggctc atttgcctctgtc acagATTCCTCCTCC GGAATCAGTGGGTC CCAGC GCGC
CTACATTGGGGT GATGATGCTGACCCTGAGTC CCAGgtatgagctttagggacagtgacatg taatgtgacc agtgtaatcagaggggggc acctc tatt
gagctttgttctcatttctgtc tttatctaagatgaaactgtgtcacacttga aataa tcaca agagc tgtctccctcatca tcttgactttctta tccc
actccac tttgtacacc tgtcaccagattgat ttcacctgt tactgctttgatttc aagcc ttcaatccat taactggca ttttaagggc catttcca
tctgtctgtaaa tcaac tttct tgt
tcttgtttatata tgggtaaatat Геном человека: 3 000 000 000 нуклеотидов gct
attgattaacactggttgta cat
tgtattttgtagtacctagccc Примерно 20 тысяч генов, < 5% генома кодирует белки ttt
ggctaataagggtgatctgtgta aag
tcacctgggctcccctgcacacgggtgagggagagggctgcagtgatgatggggatgggca aggtgtgcatgtgtc cttgaaactaggcctttgtac tcc
tcctttc tctctgtccatttttctctatagGGCTGGTCTGCGGCCTGGTGATGTGATTTTGGCCATTGGGGAGCAGATGGTACAAAATGCTGAAGATGTT
TATGAAGCTGTT CGAAC CCAATCCCAGTTGGC AGTGCAGATC CGGCGGGGAC GAGAAACACTGACCTTATATGTGAC CCCTGAGGT CACAGAATGAATAG
ATCACCAAGAGTATGAGGCTCTGCTCTGATTTCTC CTTC CTTTC TGGCTGAGGTTCTGAGGGCA CCGAGACAGAGGGTTAAATGAACCAGTGGGGC
AGGTCCCTCCAAACCACCAGCACTGACTCCTGGGCTCTGAAgAATCACAGAAA CACTTTTTATATAAAATAAAAATTATACCTAGCaacatatatagtaaa



Экзон-интронная структура и сплайсинг

- картирование генов, кодирующих белки
- альтернативный сплайсинг
- эволюция альтернативного сплайсинга
 - 2003: а.с. эволюционирует быстрее, чем конститутивный
 - 2008: положительный отбор в альт. экзонах
- количественный анализ

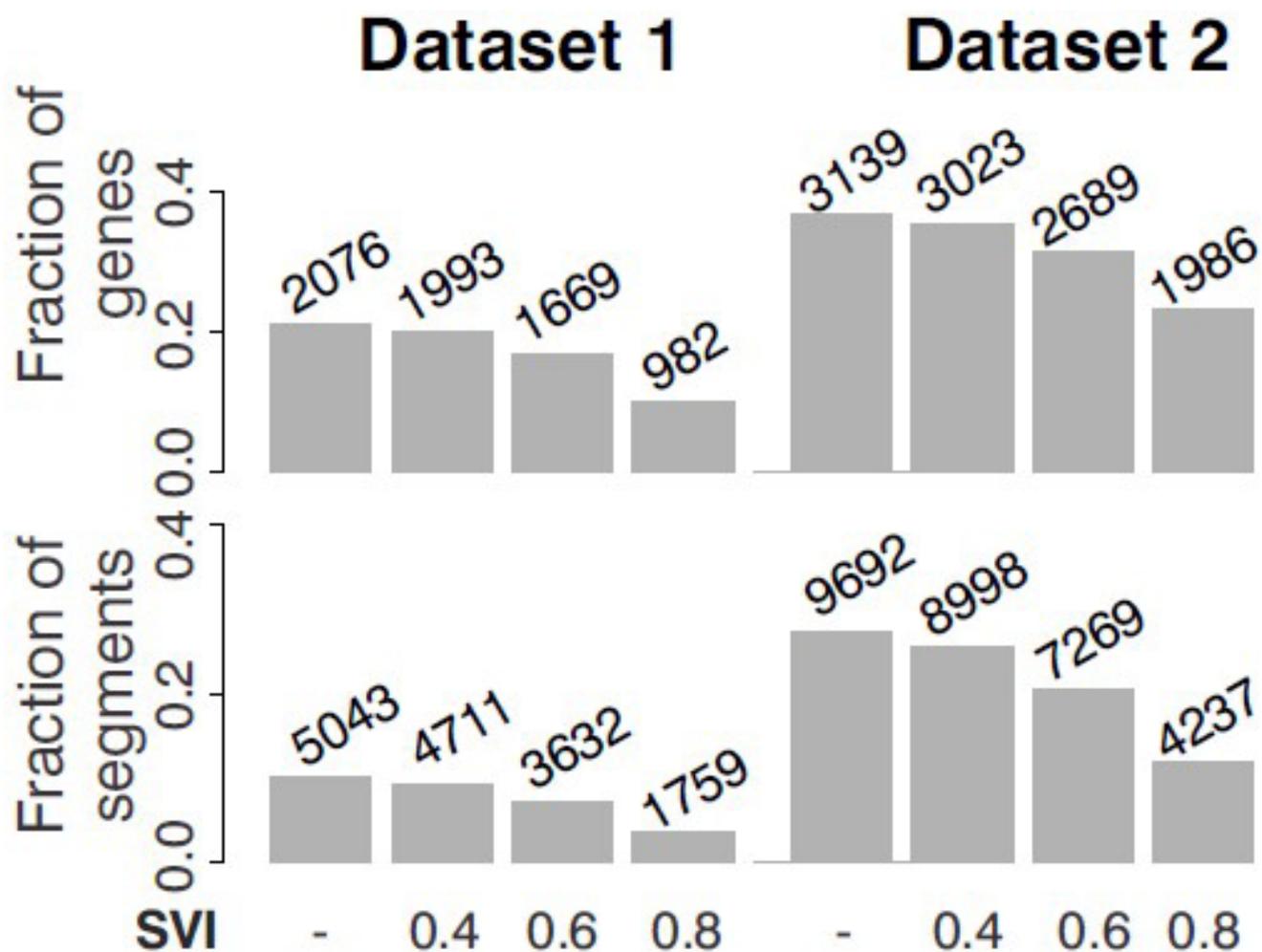
Возрастные изменения сплайсинга в мозге человека



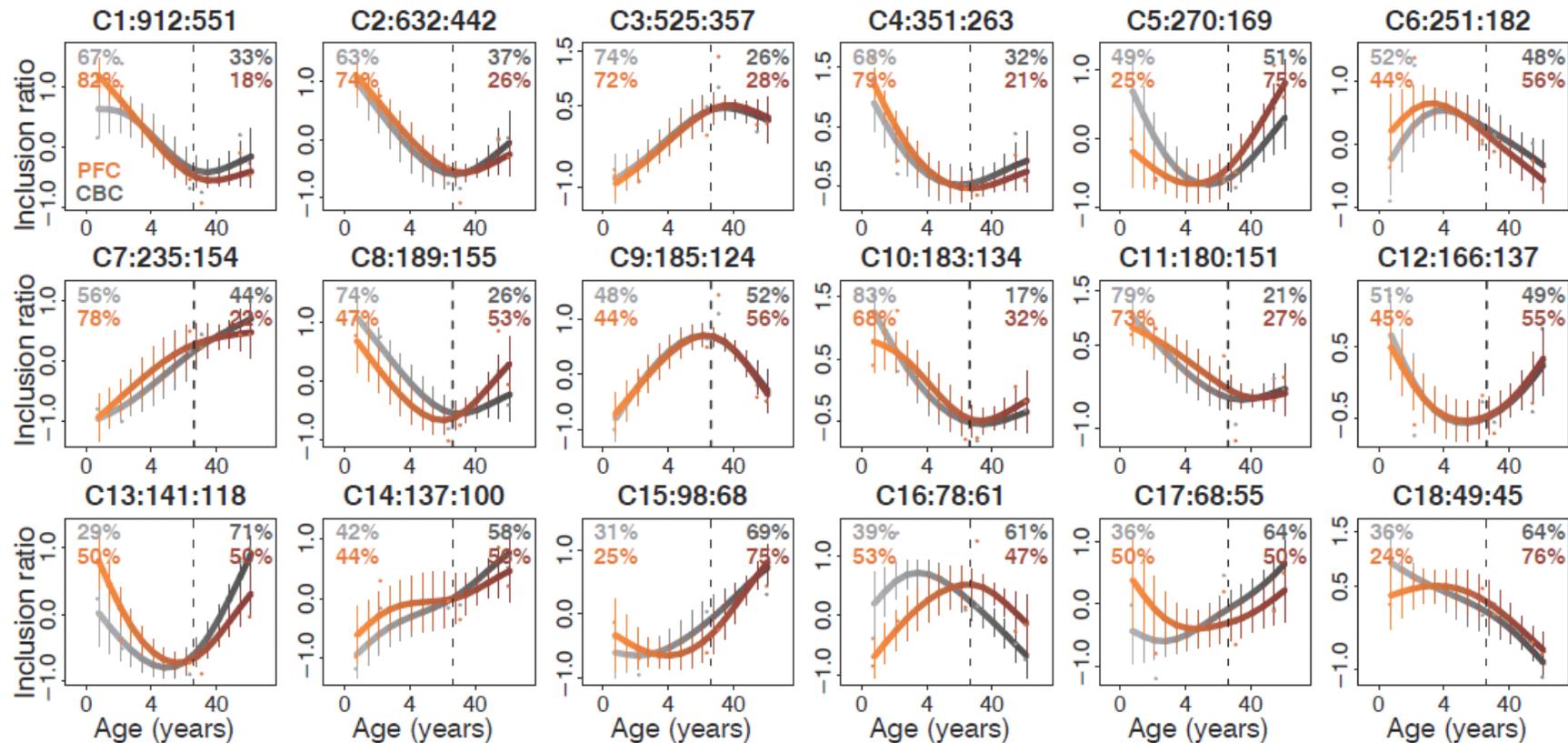
по 6 смешанных образцов (по 5 индивидуальных в каждом)

13 индивидуальных образцов

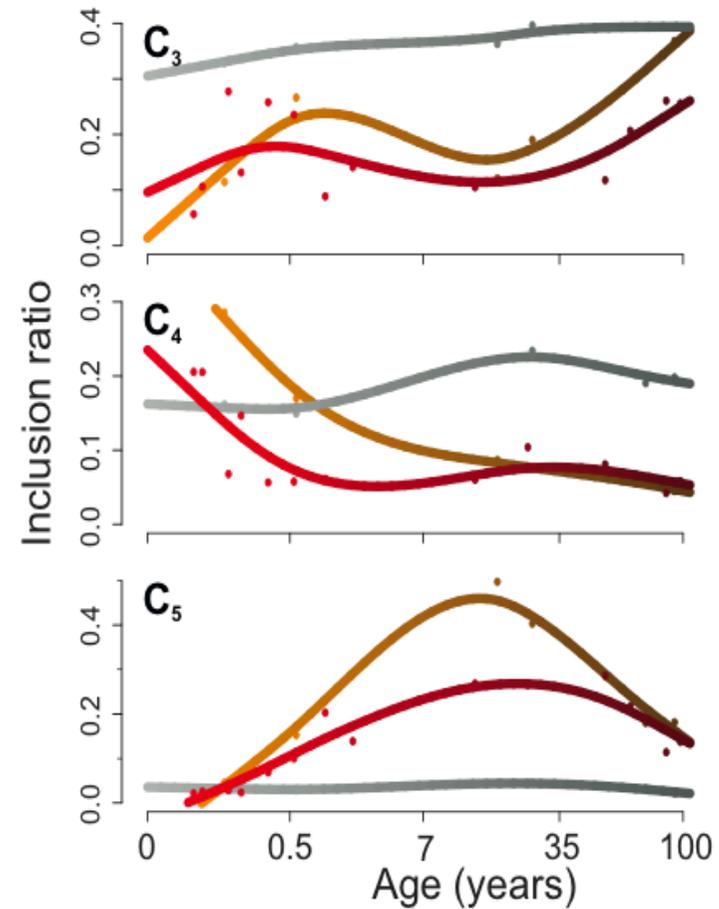
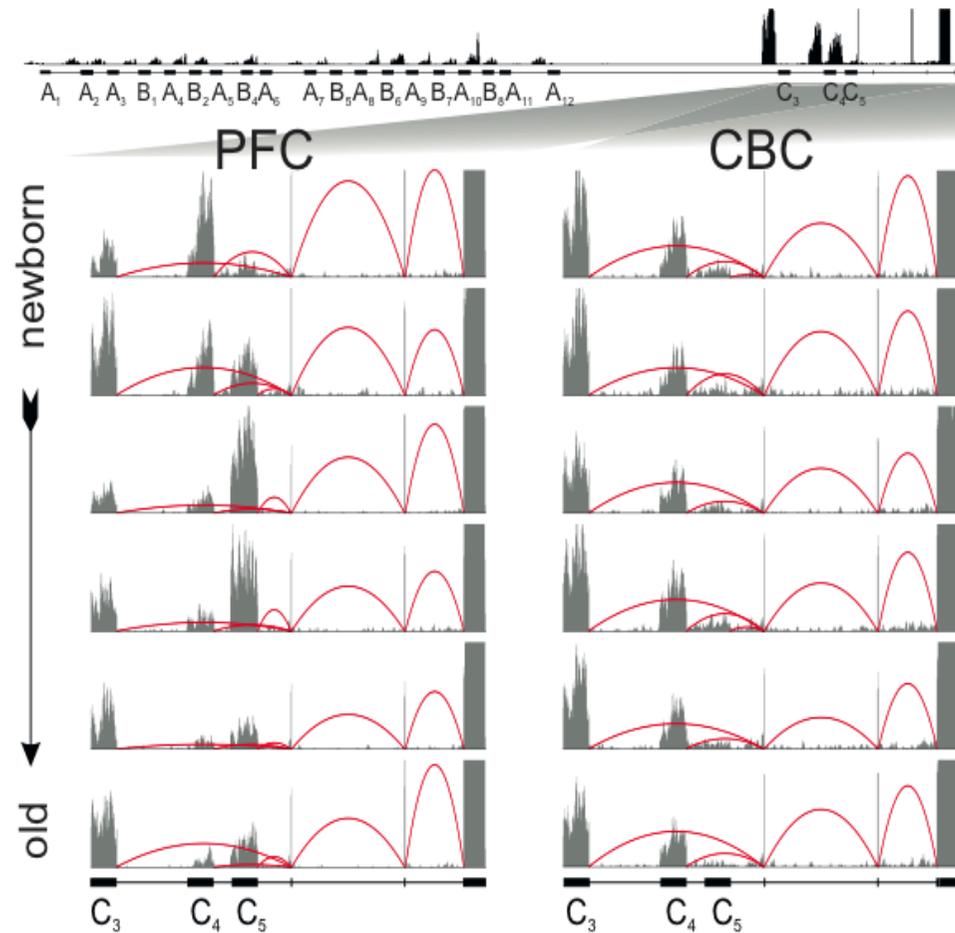
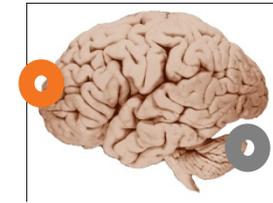
до 40% генов меняют сплайсинг с возрастом



Паттерны изменения IR с возрастом

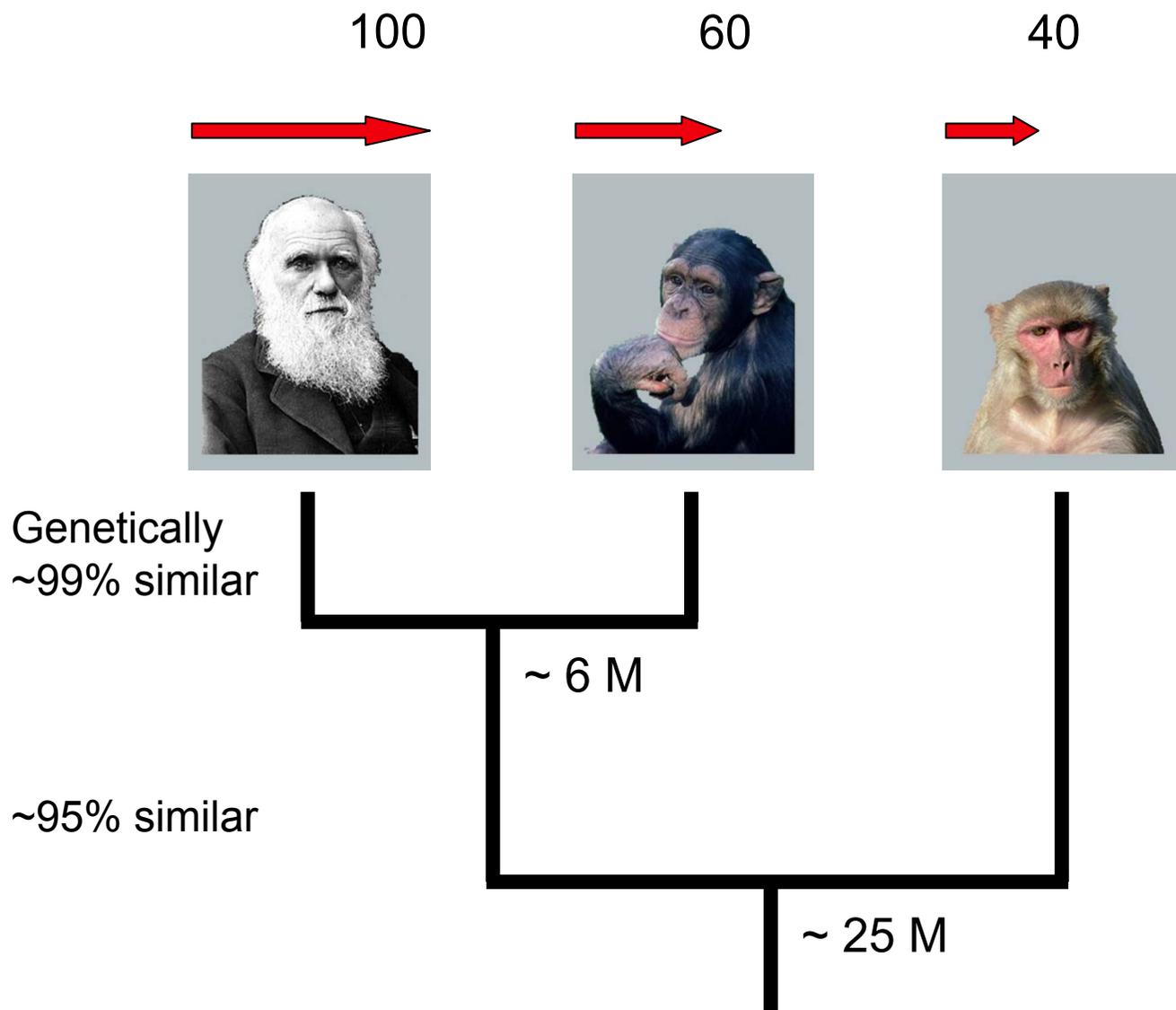


20% генов, меняющих сплайсинг с возрастом, имеют различные паттерны изменений в разных областях мозга

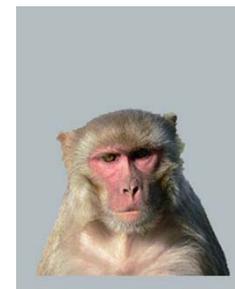
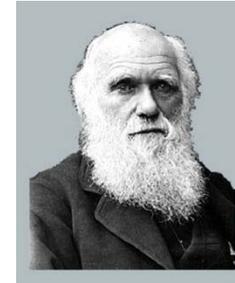
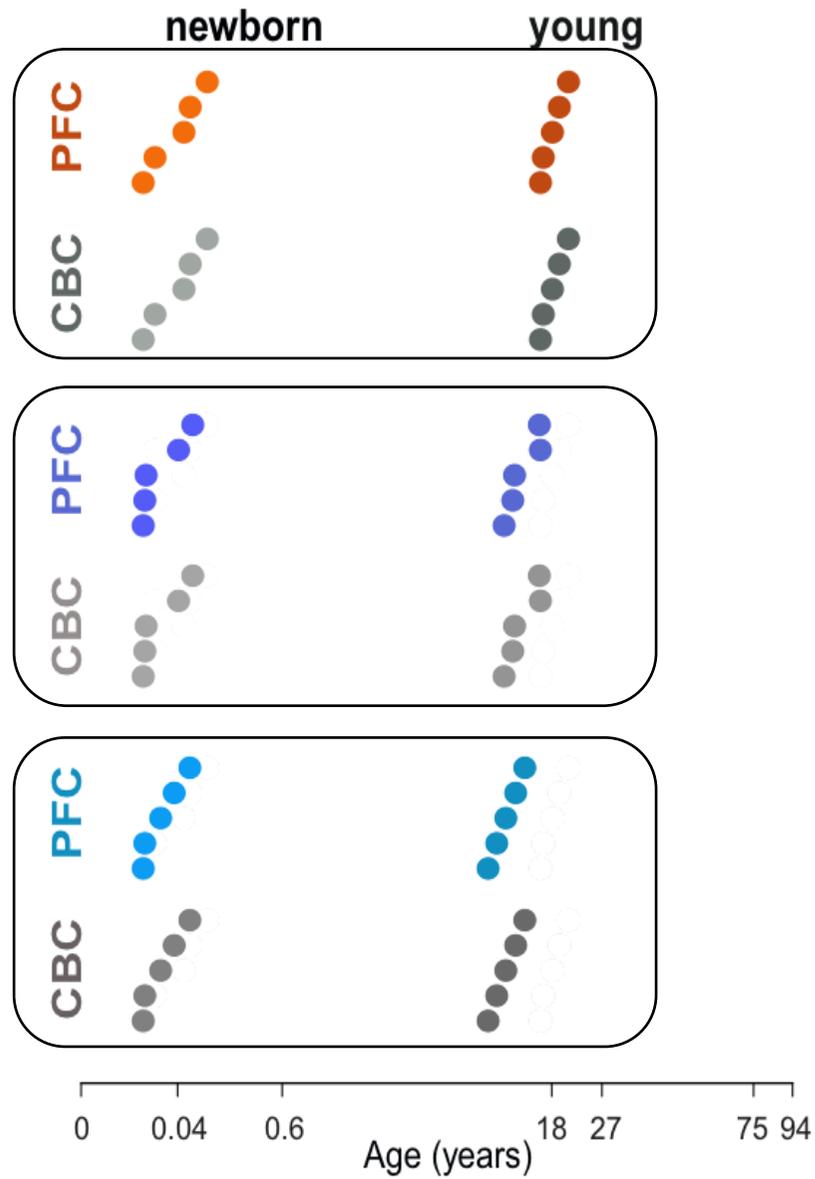


Protocadherin gamma cluster

Человек, шимпанзе, макака-резус

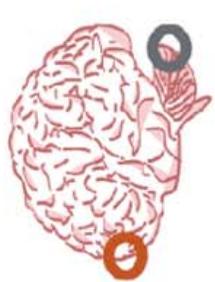


Выборка 1



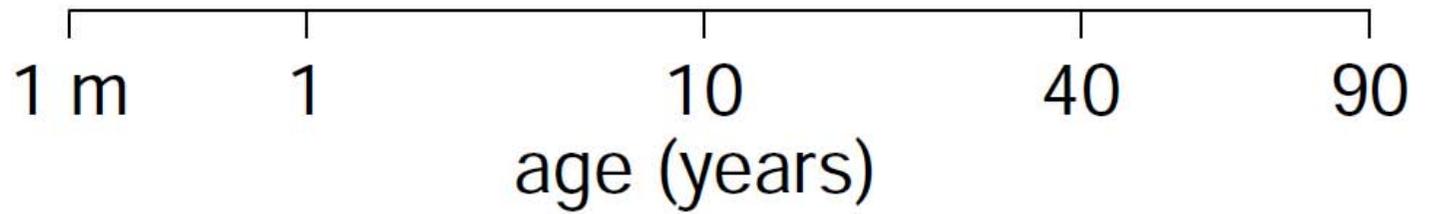
Выборки 1 и 2

Dataset 1

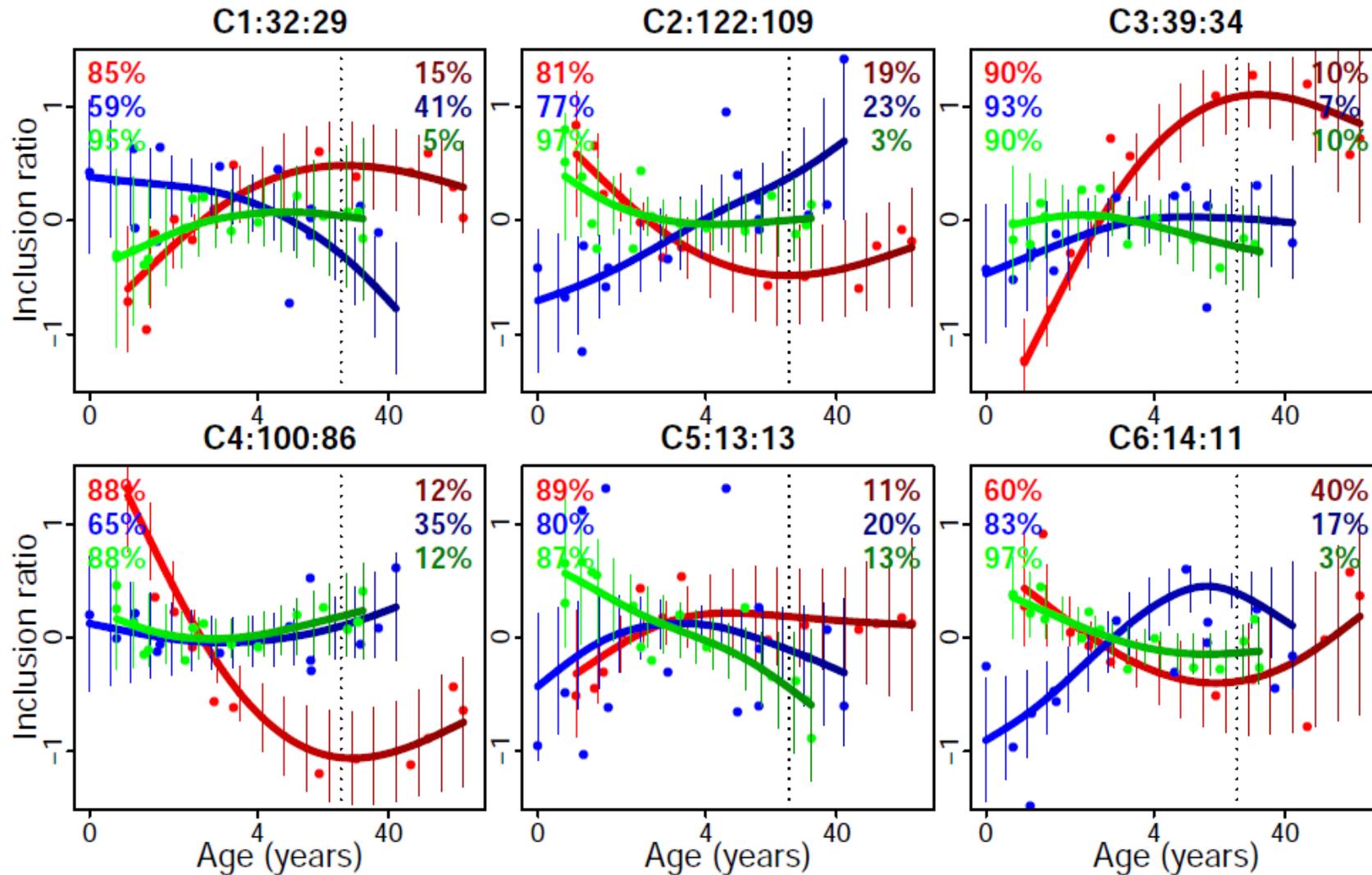


CBC
PFC

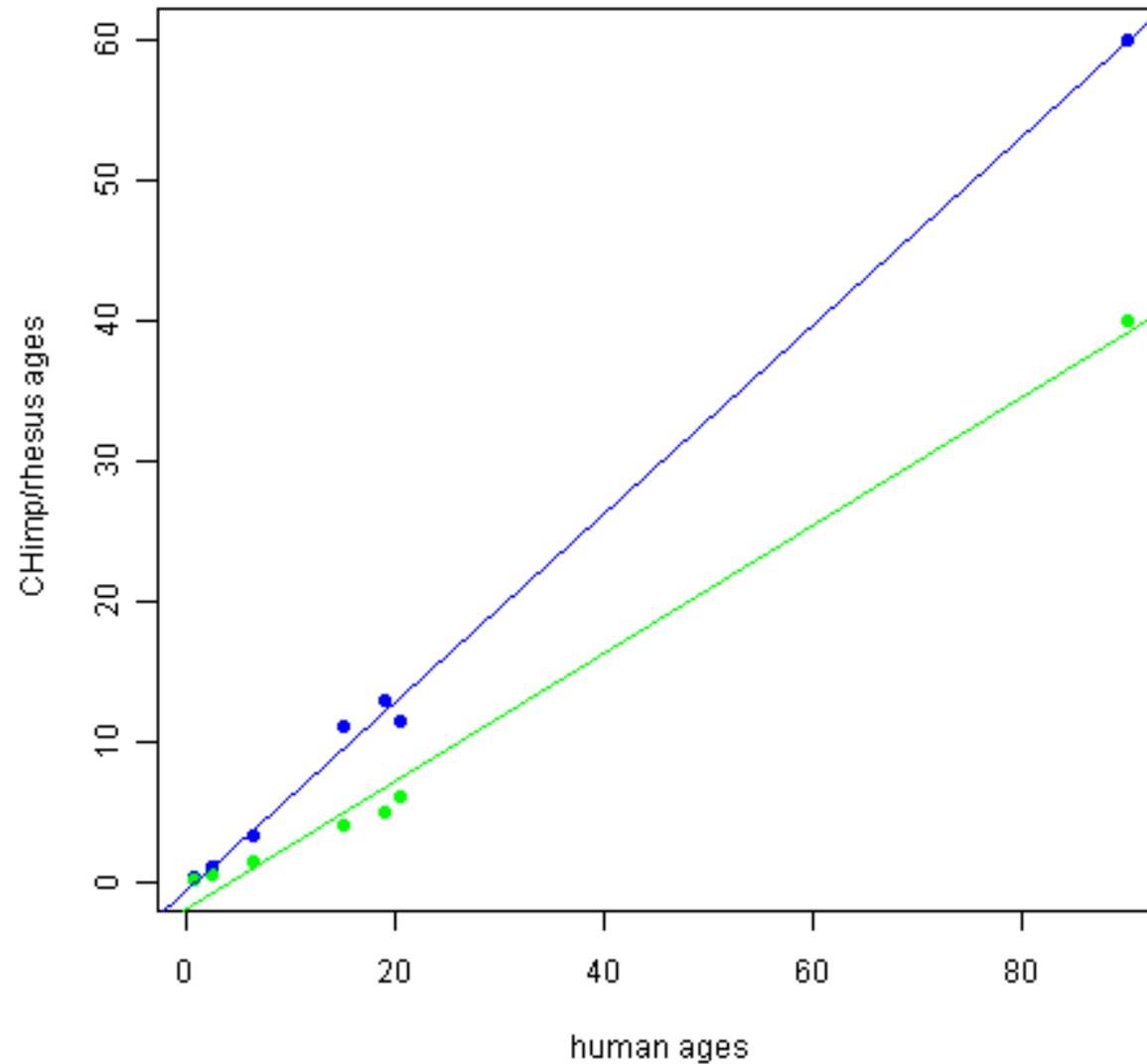
Dataset 2



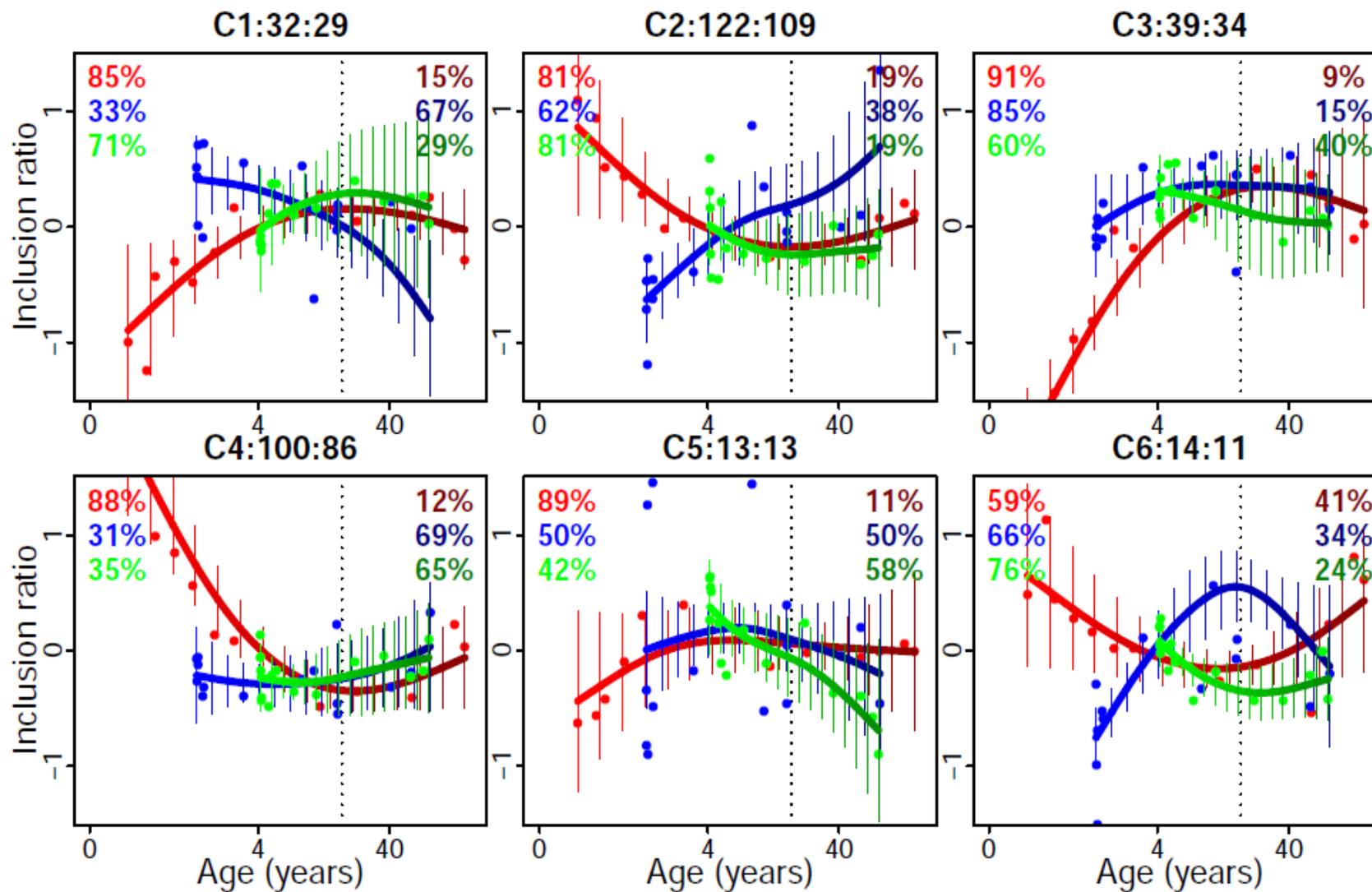
Слабое согласование профилей включения



Поправка на физиологический возраст



Теперь лучше



- Павел Мазин
- Филипп Хайтович (Шанхай)

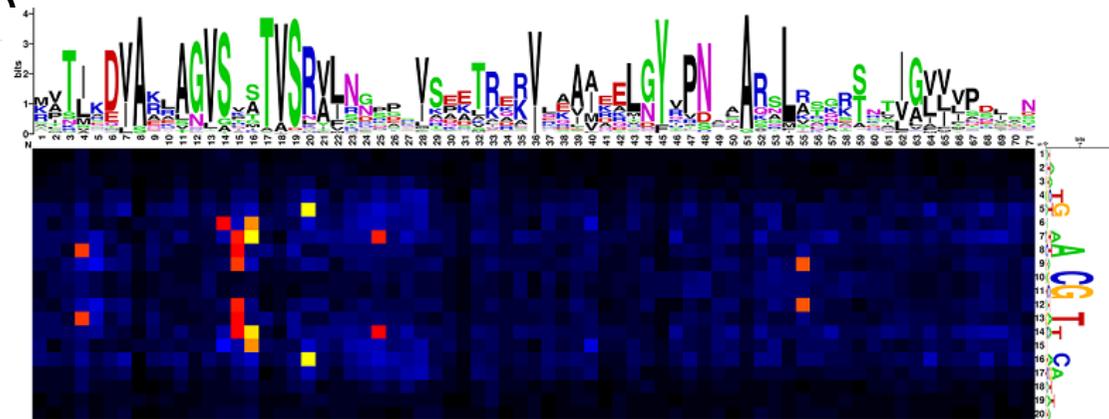


Другие проекты

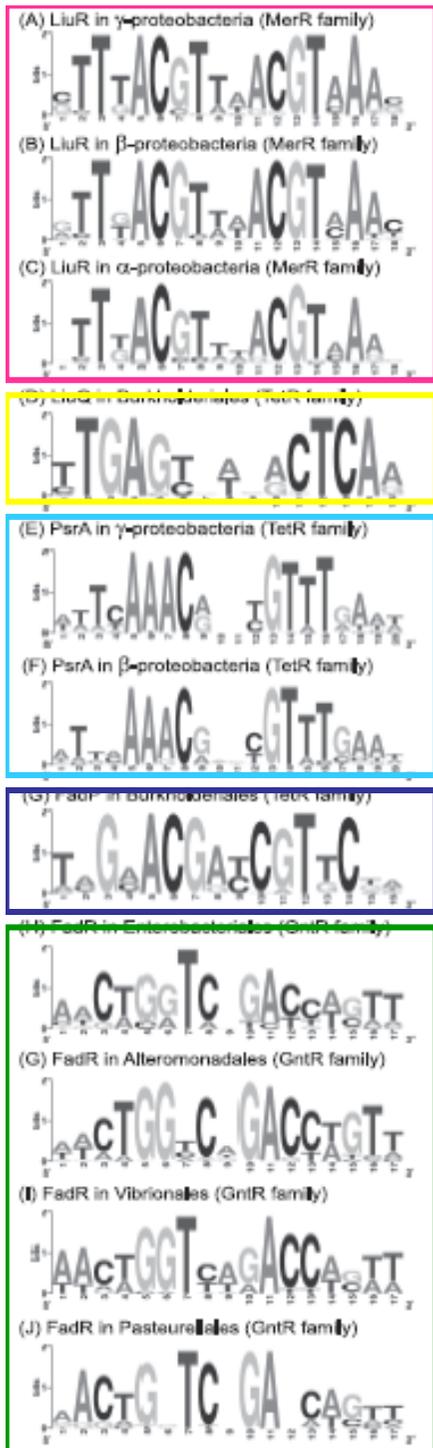
- Эволюция сплайсинга у дрозофил (Сергей Нуждин, USC + Екатерина Ермакова)
- РНК-связывающие белки (Альфред Биндерейф, Университет Гиссена + Екатерина Храмеева)
- Структура и состояние хроматина
- Старты транскрипции у бактерий (Гиссен + Институт Пастера, Елена Яловая)

Регуляторные системы

- Разработка алгоритмов (УНЦБ, Лаб. 6)
- Интернет-серверы (Инна Дубчак, LANL, Беркли)
- Сравнительная геномика и эволюция регуляторных систем
 - по функциональным системам (УНЦБ, Лаб. 6)
 - по таксонам (Д.А.Родионов, сектор 6)
 - по семействам факторов транскрипции (А.Б.Рахманинова, Юрий Коростелев)



- Моделирование регуляторных систем (В.А.Любецкий, В.И.Рубанов)



MerR

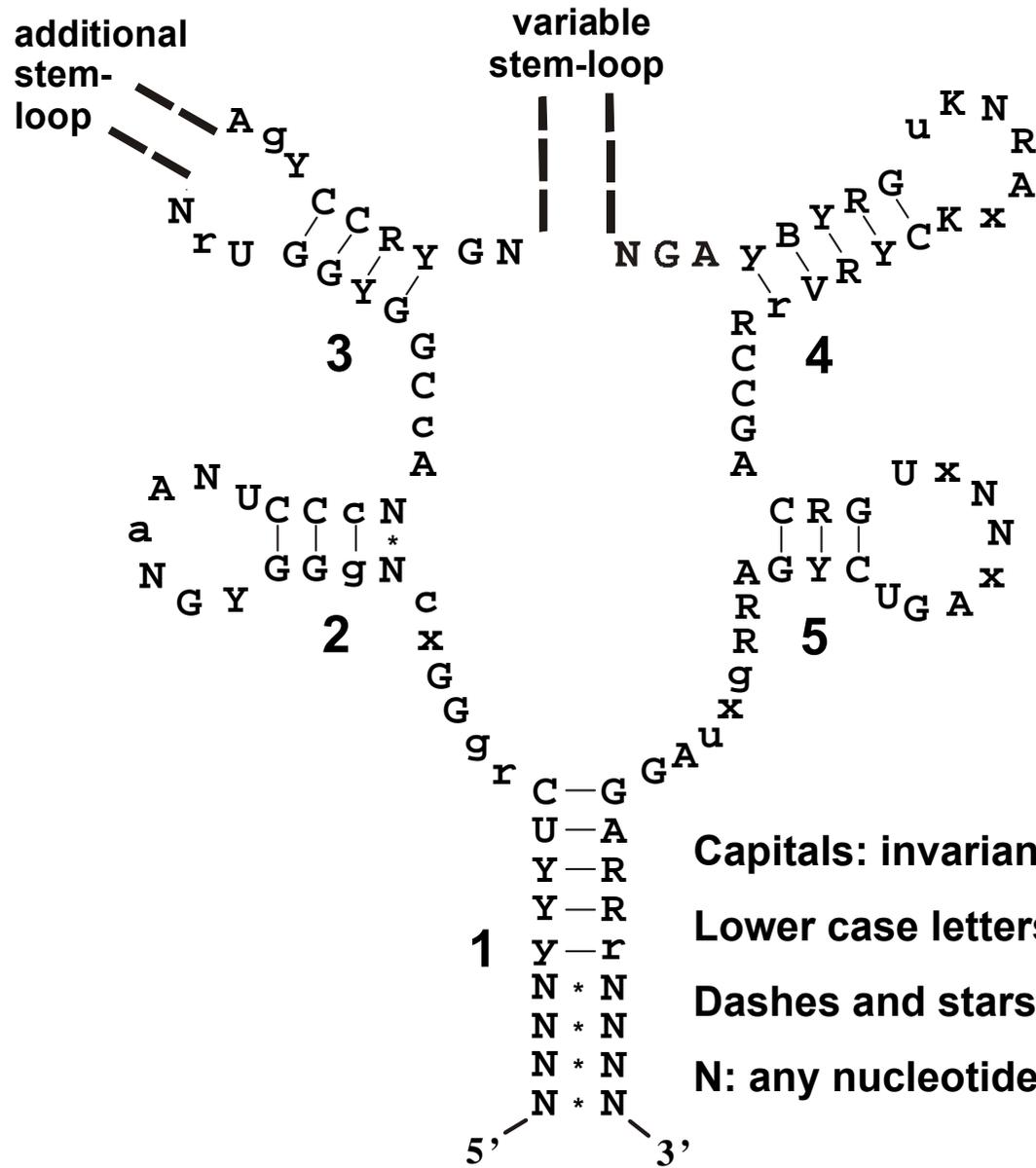
TetR

GntR

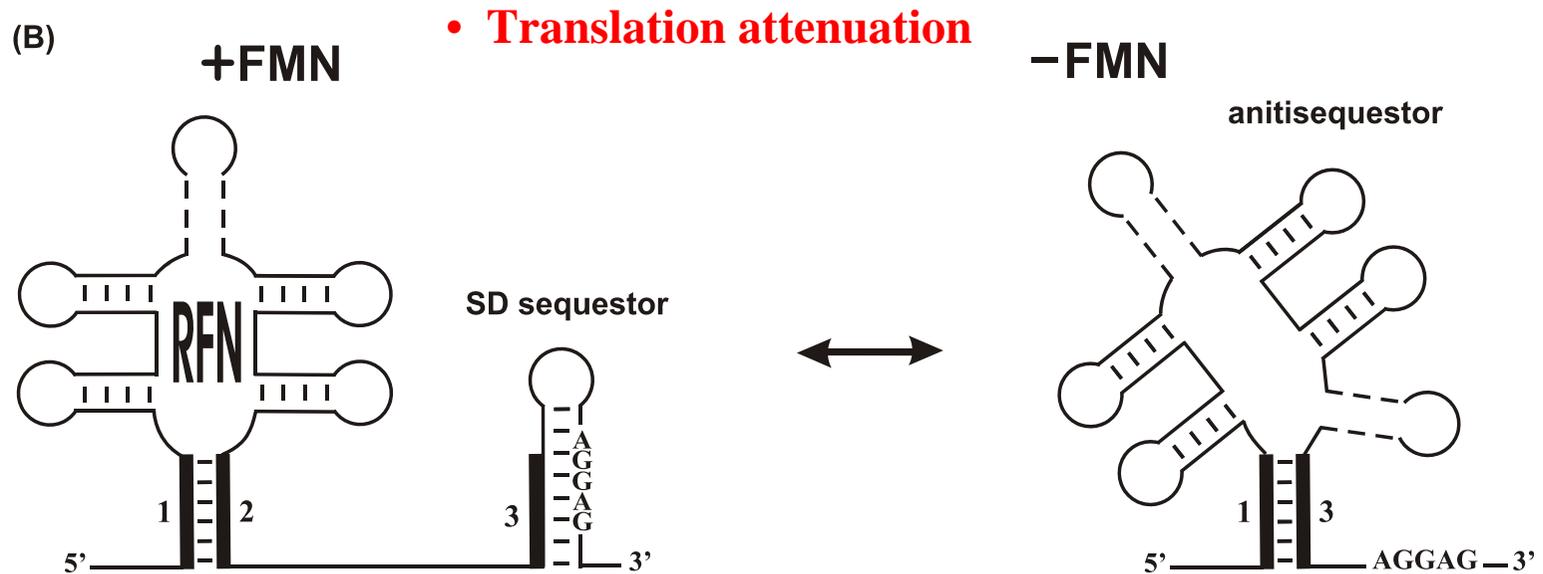
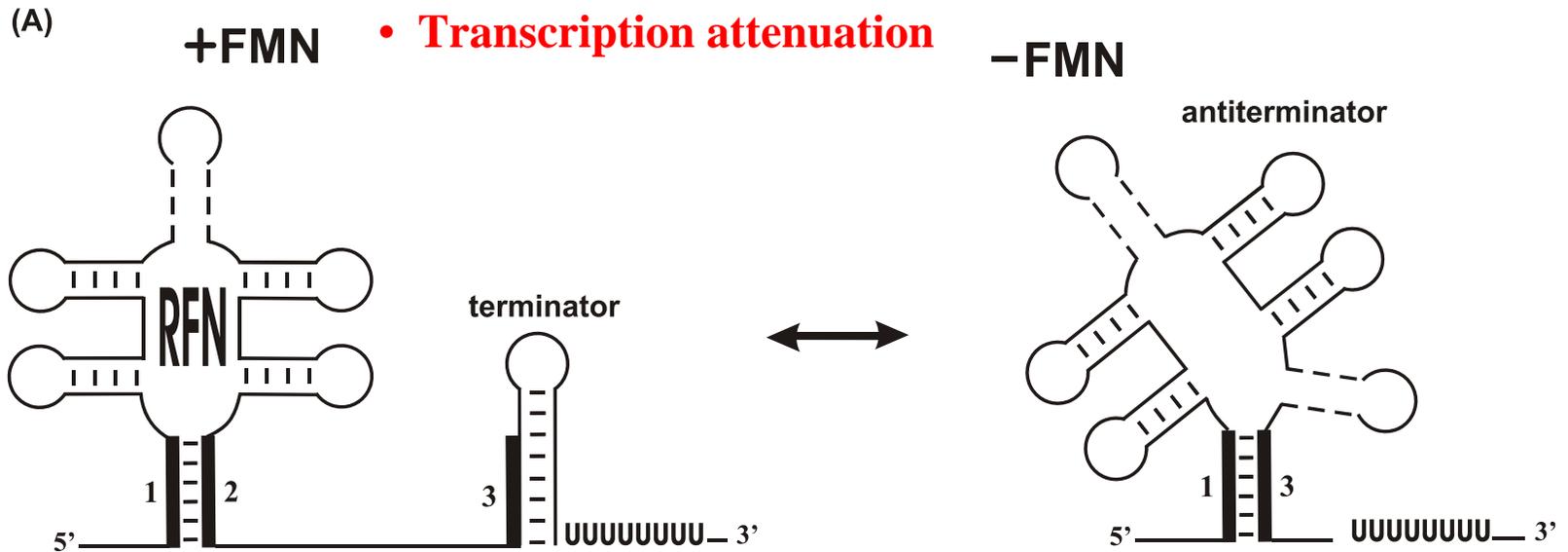
| Genome of γ -, or β -proteobacteria | ILV degradation | | | | | | | | | ETF | | FA degradation | | | | | | | | |
|---|-----------------|----------------|--------------|-------------|-------------|-------------|-------------|-----------------|------------|------------|--------------|----------------|--------------|-------------|-------------|-------------|-------------|-------------|--------------|-------------|
| | <i>liuA</i> | <i>liuBCDE</i> | <i>liuFG</i> | <i>aacS</i> | <i>ivdA</i> | <i>ivdC</i> | <i>ivdG</i> | <i>ivdIBDEF</i> | <i>bkd</i> | <i>ldh</i> | <i>etfBA</i> | <i>etfD</i> | <i>fadBA</i> | <i>fadH</i> | <i>fadD</i> | <i>fadJ</i> | <i>fadE</i> | <i>fadL</i> | <i>acdAB</i> | <i>acdH</i> |
| Enterobacteriales (5 species) | | | | | | | | | | | | | | | | | | | | |
| Pasteurellales (7 species) | | | | | | | | | | | | | | | | | | | | |
| <i>Shewanella</i> spp. | | | | | | | | | | | | | | | | | | | | |
| <i>Idiomarina loihiensis</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Colwellia psychrerythraea</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Pseudoalteromonas haloplanktis</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Pseudoalteromonas atlantica</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Saccharophagus degradans</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Vibrio cholerae</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Vibrio fischeri</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Vibrio parahaemolyticus</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Vibrio vulnificus</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Photobacterium profundum</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Pseudomonas aeruginosa</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Pseudomonas putida</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Pseudomonas fluorescens</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Pseudomonas syringae</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Pseudomonas entomophila</i> | | | | | | | | | | | | | | | | | | | | |
| Xanthomonadales (3 species) | | | | | | | | | | | | | | | | | | | | |
| <i>Hahella chejuensis</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Alcanivorax borkumensis</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Chromohalobacter salexigens</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Chromobacterium violaceum</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Dechloromonas aromatica</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Azoarcus</i> sp. | | | | | | | | | | | | | | | | | | | | |
| <i>Bordetella</i> (3 species) | | | | | | | | | | | | | | | | | | | | |
| <i>Ralstonia solanacearum</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Ralstonia eutropha</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Ralstonia metallidurans</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Burkholderia xenovorans</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Burkholderia</i> (4 species) | | | | | | | | | | | | | | | | | | | | |
| <i>Methylobium petroleiphilum</i> | | | | | | | | | | | | | | | | | | | | |
| <i>Polaromonas</i> sp. | | | | | | | | | | | | | | | | | | | | |
| <i>Rhodospirillum rubrum</i> | | | | | | | | | | | | | | | | | | | | |

■ LiuR
 ■ LiuQ
 ■ FadR
 ■ PsrA
 ■ FadP
 ■ unknown regulation

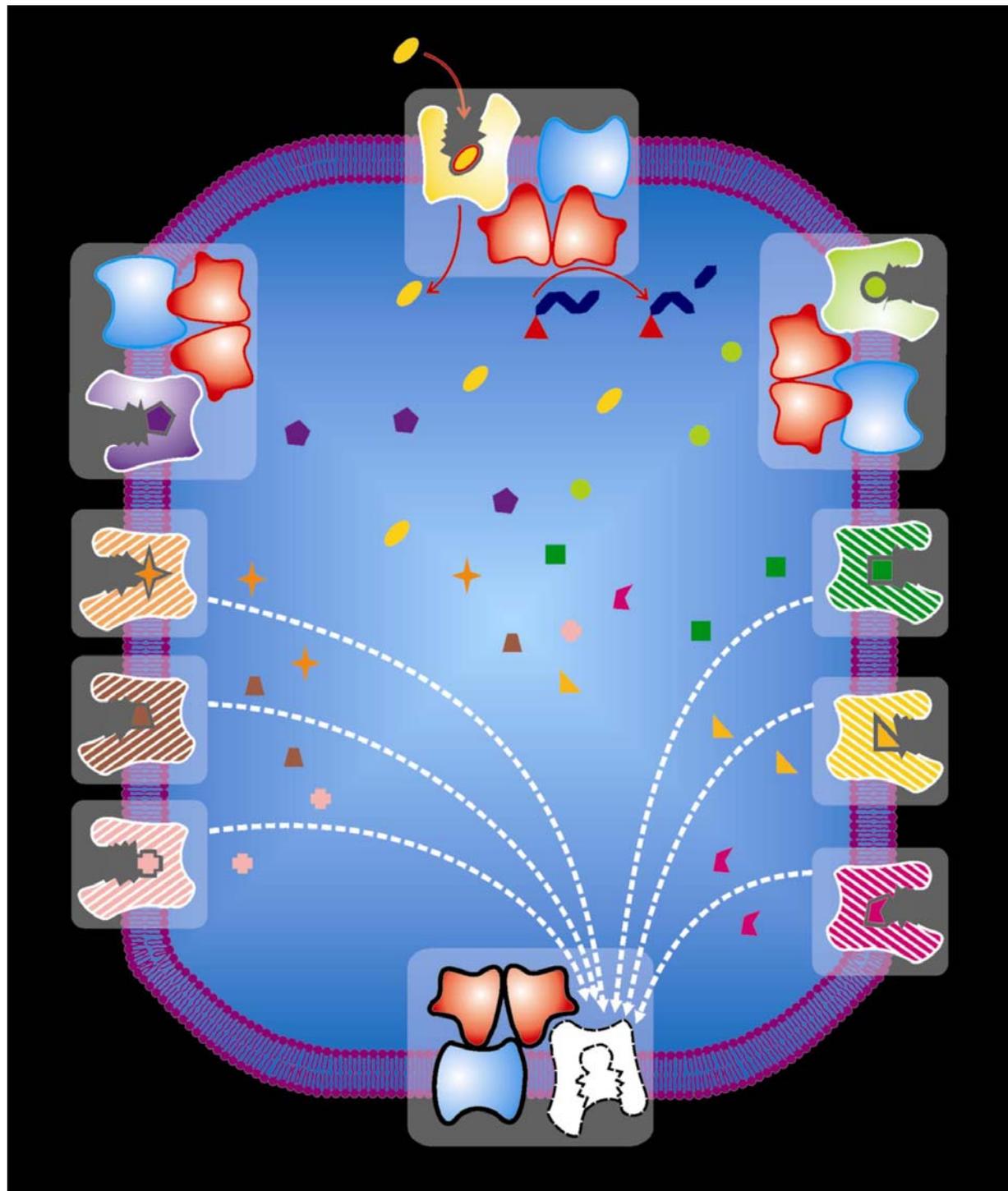
RFN - первый РНК-переключатель



RFN: механизм регуляции



Универ-
сальный
«энергети-
ческий
комплекс» +
компоненты,
определя-
ющие
специфич-
ность



Дмитрий
Родионов



Thomas
Eitingner

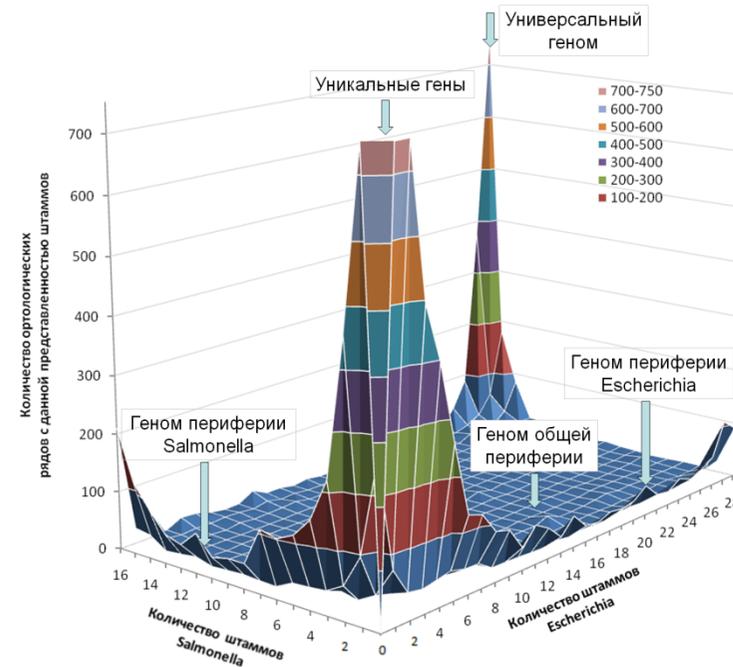
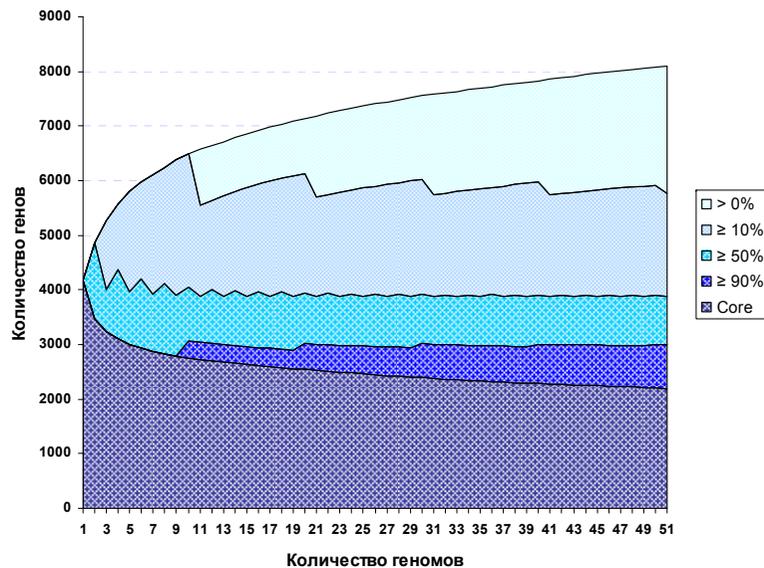


Андрей
Остерман



Эволюция бактерий

- Согласование деревьев, идентификация горизонтальных переносов (Лаб. 6)
- Микроэволюция штаммов / видов
 - на уровне наборов генов
 - геномных перестроек
 - межштаммовой рекомбинации



Лаборатория 6:

- 1) механизмы **регуляции инициации трансляции** у бактерий и в пластомах: изолейцил-тРНК синтетазы посредством Т-боксов, 2-изопропилмалатсинтазы посредством LEU-элемента у многих актиномицетов, включая возбудителей туберкулеза и дифтерии;
- 2) механизмы **регуляции транскрипции** генов метаболизма аминокислот у бактерий, метаболизма азота у цианобактерий и в хлоропластах водорослей;
- 3) механизм **процессинга мРНК** бруцелл, нарушение которого предотвращает незавершенный фагоцитоз; **регуляции транскрипции и трансляции** в пластидах споровиков – возбудителей протозойных инфекций, включая токсоплазмоз, эймериоз и малярию; предсказано воздействие антибиотиков на споровиков.
- 4) механизм **ответа изолированных хлоропластов на тепловой и холодовой шоки**; и количественные предсказания результатов **воздействия точечных мутаций и эпигенетических нарушений** в митохондриях животных на их фенотип. Объяснить влияние на фенотип генетических изменений в митохондриальной ДНК человека (**MELAS-болезни**) и нарушений метилирования ДНК при недостатке гормонов;
- 5) механизмы **задержки транскрипции до завершения сплайсинга**;
- 6) совместная **эволюция промоторов и сигма-субъединиц РНК-полимераз** пластид растений, водорослей и споровиков; **сайтов и факторов** у бактерий.





template

- text