



5 selfish reasons to work reproducibly



Florian Markowetz
CRUK Cambridge Institute
www.markowetzlab.org
 @markowetzlab



Tumour Evolution



Geoff



Ruben



Michael



Lena

Tumour Imaging



Mirela



Tristan



Paula

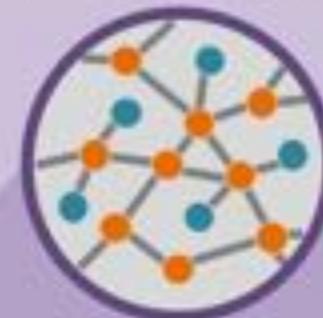


Adam



Marcel

Interaction networks



Andy



Helen



Alex



Amy



Ryan

We thank all current and past funders



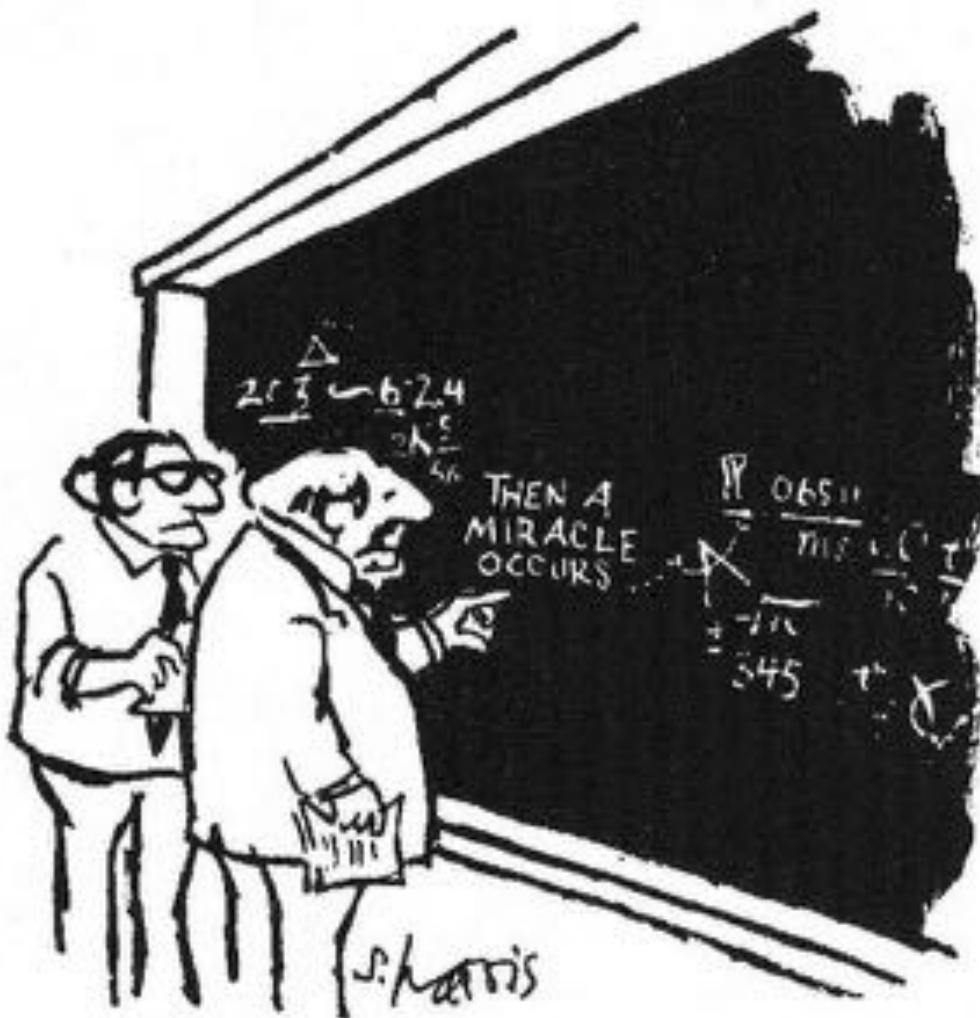
More at <http://www.markowetzlab.org>

@markowetzlab

Personal fellowships



Science ≠ miracles



"I think you should be more explicit here in step two."

“How Bright Promise in Cancer Testing Fell Apart”

New York Times 2011



```
Their List and Ours

# sort by p-value
temp <- temp[order(pValue),]
rownames(temp) <- rownames(pottiUpdated)[ftrRows]
sort(rownames(pottiUpdated) [
  ftrQNorm@p.values <= ftrCut])
> colnames(temp) <- c("Theirs", "Ours");
> temp
   Theirs      Ours
1  "1881_at"    "1882_q_at"
2  "31321_at"   "31322_at"
3  "31725_m_at" "31726_at"
4  "32307_r_at" "32308_r_at"
5  "32309_r_at" "32310_r_at"
6  "32311_r_at" "32312_r_at"
7  "32313_r_at" "32314_r_at"
8  "32315_r_at" "32316_r_at"
9  "32317_r_at" "32318_r_at"
10 "32319_r_at" "32320_r_at"
11 "32321_r_at" "32322_r_at"
12 "32323_r_at" "32324_r_at"
13 "32325_r_at" "32326_r_at"
14 "32327_r_at" "32328_r_at"
15 "32329_r_at" "32330_r_at"
16 "32331_r_at" "32332_r_at"
17 "32333_r_at" "32334_r_at"
18 "32335_r_at" "32336_r_at"
19 "32337_r_at" "32338_r_at"
20 "32339_r_at" "32340_r_at"
21 "32341_r_at" "32342_r_at"
22 "32343_r_at" "32344_r_at"
23 "32345_r_at" "32346_r_at"
24 "32347_r_at" "32348_r_at"
25 "32349_r_at" "32350_r_at"
26 "32351_r_at" "32352_r_at"
27 "32353_r_at" "32354_r_at"
28 "32355_r_at" "32356_r_at"
29 "32357_r_at" "32358_r_at"
30 "32359_r_at" "32360_r_at"
31 "32361_r_at" "32362_r_at"
32 "32363_r_at" "32364_r_at"
33 "32365_r_at" "32366_r_at"
34 "32367_r_at" "32368_r_at"
35 "32369_r_at" "32370_r_at"
36 "32371_r_at" "32372_r_at"
37 "32373_r_at" "32374_r_at"
38 "32375_r_at" "32376_r_at"
39 "32377_r_at" "32378_r_at"
40 "32379_r_at" "32380_r_at"
41 "32381_r_at" "32382_r_at"
42 "32383_r_at" "32384_r_at"
43 "32385_r_at" "32386_r_at"
44 "32387_r_at" "32388_r_at"
45 "32389_r_at" "32390_r_at"
46 "32391_r_at" "32392_r_at"
47 "32393_r_at" "32394_r_at"
48 "32395_r_at" "32396_r_at"
49 "32397_r_at" "32398_r_at"
50 "32399_r_at" "32400_r_at"
51 "32401_r_at" "32402_r_at"
52 "32403_r_at" "32404_r_at"
53 "32405_r_at" "32406_r_at"
54 "32407_r_at" "32408_r_at"
55 "32409_r_at" "32410_r_at"
56 "32411_r_at" "32412_r_at"
57 "32413_r_at" "32414_r_at"
58 "32415_r_at" "32416_r_at"
59 "32417_r_at" "32418_r_at"
60 "32419_r_at" "32420_r_at"
61 "32421_r_at" "32422_r_at"
62 "32423_r_at" "32424_r_at"
63 "32425_r_at" "32426_r_at"
64 "32427_r_at" "32428_r_at"
65 "32429_r_at" "32430_r_at"
66 "32431_r_at" "32432_r_at"
67 "32433_r_at" "32434_r_at"
68 "32435_r_at" "32436_r_at"
69 "32437_r_at" "32438_r_at"
70 "32439_r_at" "32440_r_at"
71 "32441_r_at" "32442_r_at"
72 "32443_r_at" "32444_r_at"
73 "32445_r_at" "32446_r_at"
74 "32447_r_at" "32448_r_at"
75 "32449_r_at" "32450_r_at"
76 "32451_r_at" "32452_r_at"
77 "32453_r_at" "32454_r_at"
78 "32455_r_at" "32456_r_at"
79 "32457_r_at" "32458_r_at"
80 "32459_r_at" "32460_r_at"
81 "32461_r_at" "32462_r_at"
82 "32463_r_at" "32464_r_at"
83 "32465_r_at" "32466_r_at"
84 "32467_r_at" "32468_r_at"
85 "32469_r_at" "32470_r_at"
86 "32471_r_at" "32472_r_at"
87 "32473_r_at" "32474_r_at"
88 "32475_r_at" "32476_r_at"
89 "32477_r_at" "32478_r_at"
90 "32479_r_at" "32480_r_at"
91 "32481_r_at" "32482_r_at"
92 "32483_r_at" "32484_r_at"
93 "32485_r_at" "32486_r_at"
94 "32487_r_at" "32488_r_at"
95 "32489_r_at" "32490_r_at"
96 "32491_r_at" "32492_r_at"
97 "32493_r_at" "32494_r_at"
98 "32495_r_at" "32496_r_at"
99 "32497_r_at" "32498_r_at"
100 "32499_r_at" "32500_r_at"
101 "32501_r_at" "32502_r_at"
102 "32503_r_at" "32504_r_at"
103 "32505_r_at" "32506_r_at"
104 "32507_r_at" "32508_r_at"
105 "32509_r_at" "32510_r_at"
106 "32511_r_at" "32512_r_at"
107 "32513_r_at" "32514_r_at"
108 "32515_r_at" "32516_r_at"
109 "32517_r_at" "32518_r_at"
110 "32519_r_at" "32520_r_at"
111 "32521_r_at" "32522_r_at"
112 "32523_r_at" "32524_r_at"
113 "32525_r_at" "32526_r_at"
114 "32527_r_at" "32528_r_at"
115 "32529_r_at" "32530_r_at"
116 "32531_r_at" "32532_r_at"
117 "32533_r_at" "32534_r_at"
118 "32535_r_at" "32536_r_at"
119 "32537_r_at" "32538_r_at"
120 "32539_r_at" "32540_r_at"
121 "32541_r_at" "32542_r_at"
122 "32543_r_at" "32544_r_at"
123 "32545_r_at" "32546_r_at"
124 "32547_r_at" "32548_r_at"
125 "32549_r_at" "32550_r_at"
126 "32551_r_at" "32552_r_at"
127 "32553_r_at" "32554_r_at"
128 "32555_r_at" "32556_r_at"
129 "32557_r_at" "32558_r_at"
130 "32559_r_at" "32560_r_at"
131 "32561_r_at" "32562_r_at"
132 "32563_r_at" "32564_r_at"
133 "32565_r_at" "32566_r_at"
134 "32567_r_at" "32568_r_at"
135 "32569_r_at" "32570_r_at"
136 "32571_r_at" "32572_r_at"
137 "32573_r_at" "32574_r_at"
138 "32575_r_at" "32576_r_at"
139 "32577_r_at" "32578_r_at"
140 "32579_r_at" "32580_r_at"
141 "32581_r_at" "32582_r_at"
142 "32583_r_at" "32584_r_at"
143 "32585_r_at" "32586_r_at"
144 "32587_r_at" "32588_r_at"
145 "32589_r_at" "32590_r_at"
146 "32591_r_at" "32592_r_at"
147 "32593_r_at" "32594_r_at"
148 "32595_r_at" "32596_r_at"
149 "32597_r_at" "32598_r_at"
150 "32599_r_at" "32600_r_at"
151 "32601_r_at" "32602_r_at"
152 "32603_r_at" "32604_r_at"
153 "32605_r_at" "32606_r_at"
154 "32607_r_at" "32608_r_at"
155 "32609_r_at" "32610_r_at"
156 "32611_r_at" "32612_r_at"
157 "32613_r_at" "32614_r_at"
158 "32615_r_at" "32616_r_at"
159 "32617_r_at" "32618_r_at"
160 "32619_r_at" "32620_r_at"
161 "32621_r_at" "32622_r_at"
162 "32623_r_at" "32624_r_at"
163 "32625_r_at" "32626_r_at"
164 "32627_r_at" "32628_r_at"
165 "32629_r_at" "32630_r_at"
166 "32631_r_at" "32632_r_at"
167 "32633_r_at" "32634_r_at"
168 "32635_r_at" "32636_r_at"
169 "32637_r_at" "32638_r_at"
170 "32639_r_at" "32640_r_at"
171 "32641_r_at" "32642_r_at"
172 "32643_r_at" "32644_r_at"
173 "32645_r_at" "32646_r_at"
174 "32647_r_at" "32648_r_at"
175 "32649_r_at" "32650_r_at"
176 "32651_r_at" "32652_r_at"
177 "32653_r_at" "32654_r_at"
178 "32655_r_at" "32656_r_at"
179 "32657_r_at" "32658_r_at"
180 "32659_r_at" "32660_r_at"
181 "32661_r_at" "32662_r_at"
182 "32663_r_at" "32664_r_at"
183 "32665_r_at" "32666_r_at"
184 "32667_r_at" "32668_r_at"
185 "32669_r_at" "32670_r_at"
186 "32671_r_at" "32672_r_at"
187 "32673_r_at" "32674_r_at"
188 "32675_r_at" "32676_r_at"
189 "32677_r_at" "32678_r_at"
190 "32679_r_at" "32680_r_at"
191 "32681_r_at" "32682_r_at"
192 "32683_r_at" "32684_r_at"
193 "32685_r_at" "32686_r_at"
194 "32687_r_at" "32688_r_at"
195 "32689_r_at" "32690_r_at"
196 "32691_r_at" "32692_r_at"
197 "32693_r_at" "32694_r_at"
198 "32695_r_at" "32696_r_at"
199 "32697_r_at" "32698_r_at"
200 "32699_r_at" "32700_r_at"
201 "32701_r_at" "32702_r_at"
202 "32703_r_at" "32704_r_at"
203 "32705_r_at" "32706_r_at"
204 "32707_r_at" "32708_r_at"
205 "32709_r_at" "32710_r_at"
206 "32711_r_at" "32712_r_at"
207 "32713_r_at" "32714_r_at"
208 "32715_r_at" "32716_r_at"
209 "32717_r_at" "32718_r_at"
210 "32719_r_at" "32720_r_at"
211 "32721_r_at" "32722_r_at"
212 "32723_r_at" "32724_r_at"
213 "32725_r_at" "32726_r_at"
214 "32727_r_at" "32728_r_at"
215 "32729_r_at" "32730_r_at"
216 "32731_r_at" "32732_r_at"
217 "32733_r_at" "32734_r_at"
218 "32735_r_at" "32736_r_at"
219 "32737_r_at" "32738_r_at"
220 "32739_r_at" "32740_r_at"
221 "32741_r_at" "32742_r_at"
222 "32743_r_at" "32744_r_at"
223 "32745_r_at" "32746_r_at"
224 "32747_r_at" "32748_r_at"
225 "32749_r_at" "32750_r_at"
226 "32751_r_at" "32752_r_at"
227 "32753_r_at" "32754_r_at"
228 "32755_r_at" "32756_r_at"
229 "32757_r_at" "32758_r_at"
230 "32759_r_at" "32760_r_at"
231 "32761_r_at" "32762_r_at"
232 "32763_r_at" "32764_r_at"
233 "32765_r_at" "32766_r_at"
234 "32767_r_at" "32768_r_at"
235 "32769_r_at" "32770_r_at"
236 "32771_r_at" "32772_r_at"
237 "32773_r_at" "32774_r_at"
238 "32775_r_at" "32776_r_at"
239 "32777_r_at" "32778_r_at"
240 "32779_r_at" "32780_r_at"
241 "32781_r_at" "32782_r_at"
242 "32783_r_at" "32784_r_at"
243 "32785_r_at" "32786_r_at"
244 "32787_r_at" "32788_r_at"
245 "32789_r_at" "32790_r_at"
246 "32791_r_at" "32792_r_at"
247 "32793_r_at" "32794_r_at"
248 "32795_r_at" "32796_r_at"
249 "32797_r_at" "32798_r_at"
250 "32799_r_at" "32800_r_at"
251 "32801_r_at" "32802_r_at"
252 "32803_r_at" "32804_r_at"
253 "32805_r_at" "32806_r_at"
254 "32807_r_at" "32808_r_at"
255 "32809_r_at" "32810_r_at"
256 "32811_r_at" "32812_r_at"
257 "32813_r_at" "32814_r_at"
258 "32815_r_at" "32816_r_at"
259 "32817_r_at" "32818_r_at"
260 "32819_r_at" "32820_r_at"
261 "32821_r_at" "32822_r_at"
262 "32823_r_at" "32824_r_at"
263 "32825_r_at" "32826_r_at"
264 "32827_r_at" "32828_r_at"
265 "32829_r_at" "32830_r_at"
266 "32831_r_at" "32832_r_at"
267 "32833_r_at" "32834_r_at"
268 "32835_r_at" "32836_r_at"
269 "32837_r_at" "32838_r_at"
270 "32839_r_at" "32840_r_at"
271 "32841_r_at" "32842_r_at"
272 "32843_r_at" "32844_r_at"
273 "32845_r_at" "32846_r_at"
274 "32847_r_at" "32848_r_at"
275 "32849_r_at" "32850_r_at"
276 "32851_r_at" "32852_r_at"
277 "32853_r_at" "32854_r_at"
278 "32855_r_at" "32856_r_at"
279 "32857_r_at" "32858_r_at"
280 "32859_r_at" "32860_r_at"
281 "32861_r_at" "32862_r_at"
282 "32863_r_at" "32864_r_at"
283 "32865_r_at" "32866_r_at"
284 "32867_r_at" "32868_r_at"
285 "32869_r_at" "32870_r_at"
286 "32871_r_at" "32872_r_at"
287 "32873_r_at" "32874_r_at"
288 "32875_r_at" "32876_r_at"
289 "32877_r_at" "32878_r_at"
290 "32879_r_at" "32880_r_at"
291 "32881_r_at" "32882_r_at"
292 "32883_r_at" "32884_r_at"
293 "32885_r_at" "32886_r_at"
294 "32887_r_at" "32888_r_at"
295 "32889_r_at" "32890_r_at"
296 "32891_r_at" "32892_r_at"
297 "32893_r_at" "32894_r_at"
298 "32895_r_at" "32896_r_at"
299 "32897_r_at" "32898_r_at"
300 "32899_r_at" "32900_r_at"
301 "32901_r_at" "32902_r_at"
302 "32903_r_at" "32904_r_at"
303 "32905_r_at" "32906_r_at"
304 "32907_r_at" "32908_r_at"
305 "32909_r_at" "32910_r_at"
306 "32911_r_at" "32912_r_at"
307 "32913_r_at" "32914_r_at"
308 "32915_r_at" "32916_r_at"
309 "32917_r_at" "32918_r_at"
310 "32919_r_at" "32920_r_at"
311 "32921_r_at" "32922_r_at"
312 "32923_r_at" "32924_r_at"
313 "32925_r_at" "32926_r_at"
314 "32927_r_at" "32928_r_at"
315 "32929_r_at" "32930_r_at"
316 "32931_r_at" "32932_r_at"
317 "32933_r_at" "32934_r_at"
318 "32935_r_at" "32936_r_at"
319 "32937_r_at" "32938_r_at"
320 "32939_r_at" "32940_r_at"
321 "32941_r_at" "32942_r_at"
322 "32943_r_at" "32944_r_at"
323 "32945_r_at" "32946_r_at"
324 "32947_r_at" "32948_r_at"
325 "32949_r_at" "32950_r_at"
326 "32951_r_at" "32952_r_at"
327 "32953_r_at" "32954_r_at"
328 "32955_r_at" "32956_r_at"
329 "32957_r_at" "32958_r_at"
330 "32959_r_at" "32960_r_at"
331 "32961_r_at" "32962_r_at"
332 "32963_r_at" "32964_r_at"
333 "32965_r_at" "32966_r_at"
334 "32967_r_at" "32968_r_at"
335 "32969_r_at" "32970_r_at"
336 "32971_r_at" "32972_r_at"
337 "32973_r_at" "32974_r_at"
338 "32975_r_at" "32976_r_at"
339 "32977_r_at" "32978_r_at"
340 "32979_r_at" "32980_r_at"
341 "32981_r_at" "32982_r_at"
342 "32983_r_at" "32984_r_at"
343 "32985_r_at" "32986_r_at"
344 "32987_r_at" "32988_r_at"
345 "32989_r_at" "32990_r_at"
346 "32991_r_at" "32992_r_at"
347 "32993_r_at" "32994_r_at"
348 "32995_r_at" "32996_r_at"
349 "32997_r_at" "32998_r_at"
350 "32999_r_at" "33000_r_at"
351 "33001_r_at" "33002_r_at"
352 "33003_r_at" "33004_r_at"
353 "33005_r_at" "33006_r_at"
354 "33007_r_at" "33008_r_at"
355 "33009_r_at" "33010_r_at"
356 "33011_r_at" "33012_r_at"
357 "33013_r_at" "33014_r_at"
358 "33015_r_at" "33016_r_at"
359 "33017_r_at" "33018_r_at"
360 "33019_r_at" "33020_r_at"
361 "33021_r_at" "33022_r_at"
362 "33023_r_at" "33024_r_at"
363 "33025_r_at" "33026_r_at"
364 "33027_r_at" "33028_r_at"
365 "33029_r_at" "33030_r_at"
366 "33031_r_at" "33032_r_at"
367 "33033_r_at" "33034_r_at"
368 "33035_r_at" "33036_r_at"
369 "33037_r_at" "33038_r_at"
370 "33039_r_at" "33040_r_at"
371 "33041_r_at" "33042_r_at"
372 "33043_r_at" "33044_r_at"
373 "33045_r_at" "33046_r_at"
374 "33047_r_at" "33048_r_at"
375 "33049_r_at" "33050_r_at"
376 "33051_r_at" "33052_r_at"
377 "33053_r_at" "33054_r_at"
378 "33055_r_at" "33056_r_at"
379 "33057_r_at" "33058_r_at"
380 "33059_r_at" "33060_r_at"
381 "33061_r_at" "33062_r_at"
382 "33063_r_at" "33064_r_at"
383 "33065_r_at" "33066_r_at"
384 "33067_r_at" "33068_r_at"
385 "33069_r_at" "33070_r_at"
386 "33071_r_at" "33072_r_at"
387 "33073_r_at" "33074_r_at"
388 "33075_r_at" "33076_r_at"
389 "33077_r_at" "33078_r_at"
390 "33079_r_at" "33080_r_at"
391 "33081_r_at" "33082_r_at"
392 "33083_r_at" "33084_r_at"
393 "33085_r_at" "33086_r_at"
394 "33087_r_at" "33088_r_at"
395 "33089_r_at" "33090_r_at"
396 "33091_r_at" "33092_r_at"
397 "33093_r_at" "33094_r_at"
398 "33095_r_at" "33096_r_at"
399 "33097_r_at" "33098_r_at"
400 "33099_r_at" "33100_r_at"
401 "33101_r_at" "33102_r_at"
402 "33103_r_at" "33104_r_at"
403 "33105_r_at" "33106_r_at"
404 "33107_r_at" "33108_r_at"
405 "33109_r_at" "33110_r_at"
406 "33111_r_at" "33112_r_at"
407 "33113_r_at" "33114_r_at"
408 "33115_r_at" "33116_r_at"
409 "33117_r_at" "33118_r_at"
410 "33119_r_at" "33120_r_at"
411 "33121_r_at" "33122_r_at"
412 "33123_r_at" "33124_r_at"
413 "33125_r_at" "33126_r_at"
414 "33127_r_at" "33128_r_at"
415 "33129_r_at" "33130_r_at"
416 "33131_r_at" "33132_r_at"
417 "33133_r_at" "33134_r_at"
418 "33135_r_at" "33136_r_at"
419 "33137_r_at" "33138_r_at"
420 "33139_r_at" "33140_r_at"
421 "33141_r_at" "33142_r_at"
422 "33143_r_at" "33144_r_at"
423 "33145_r_at" "33146_r_at"
424 "33147_r_at" "33148_r_at"
425 "33149_r_at" "33150_r_at"
426 "33151_r_at" "33152_r_at"
427 "33153_r_at" "33154_r_at"
428 "33155_r_at" "33156_r_at"
429 "33157_r_at" "33158_r_at"
430 "33159_r_at" "33160_r_at"
431 "33161_r_at" "33162_r_at"
432 "33163_r_at" "33164_r_at"
433 "33165_r_at" "33166_r_at"
434 "33167_r_at" "33168_r_at"
435 "33169_r_at" "33170_r_at"
436 "33171_r_at" "33172_r_at"
437 "33173_r_at" "33174_r_at"
438 "33175_r_at" "33176_r_at"
439 "33177_r_at" "33178_r_at"
440 "33179_r_at" "33180_r_at"
441 "33181_r_at" "33182_r_at"
442 "33183_r_at" "33184_r_at"
443 "33185_r_at" "33186_r_at"
444 "33187_r_at" "33188_r_at"
445 "33189_r_at" "33190_r_at"
446 "33191_r_at" "33192_r_at"
447 "33193_r_at" "33194_r_at"
448 "33195_r_at" "33196_r_at"
449 "33197_r_at" "33198_r_at"
450 "33199_r_at" "33200_r_at"
451 "33201_r_at" "33202_r_at"
452 "33203_r_at" "33204_r_at"
453 "33205_r_at" "33206_r_at"
454 "33207_r_at" "33208_r_at"
455 "33209_r_at" "33210_r_at"
456 "33211_r_at" "33212_r_at"
457 "33213_r_at" "33214_r_at"
458 "33215_r_at" "33216_r_at"
459 "33217_r_at" "33218_r_at"
460 "33219_r_at" "33220_r_at"
461 "33221_r_at" "33222_r_at"
462 "33223_r_at" "33224_r_at"
463 "33225_r_at" "33226_r_at"
464 "33227_r_at" "33228_r_at"
465 "33229_r_at" "33230_r_at"
466 "33231_r_at" "33232_r_at"
467 "33233_r_at" "33234_r_at"
468 "33235_r_at" "33236_r_at"
469 "33237_r_at" "33238_r_at"
470 "33239_r_at" "33240_r_at"
471 "33241_r_at" "33242_r_at"
472 "33243_r_at" "33244_r_at"
473 "33245_r_at" "33246_r_at"
474 "33247_r_at" "33248_r_at"
475 "33249_r_at" "33250_r_at"
476 "33251_r_at" "33252_r_at"
477 "33253_r_at" "33254_r_at"
478 "33255_r_at" "33256_r_at"
479 "33257_r_at" "33258_r_at"
480 "33259_r_at" "33260_r_at"
481 "33261_r_at" "33262_r_at"
482 "33263_r_at" "33264_r_at"
483 "33265_r_at" "33266_r_at"
484 "33267_r_at" "33268_r_at"
485 "33269_r_at" "33270_r_at"
486 "33271_r_at" "33272_r_at"
487 "33273_r_at" "33274_r_at"
488 "33275_r_at" "33276_r_at"
489 "33277_r_at" "33278_r_at"
490 "33279_r_at" "33280_r_at"
491 "33281_r_at" "33282_r_at"
492 "33283_r_at" "33284_r_at"
493 "33285_r_at" "33286_r_at"
494 "33287_r_at" "33288_r_at"
495 "33289_r_at" "33290_r_at"
496 "33291_r_at" "33292_r_at"
497 "33293_r_at" "33294_r_at"
498 "33295_r_at" "33296_r_at"
499 "33297_r_at" "33298_r_at"
500 "33299_r_at" "33300_r_at"
501 "33301_r_at" "33302_r_at"
502 "33303_r_at" "33304_r_at"
503 "33305_r_at" "33306_r_at"
504 "33307_r_at" "33308_r_at"
505 "33309_r_at" "33310_r_at"
506 "33311_r_at" "33312_r_at"
507 "33313_r_at" "33314_r_at"
508 "33315_r_at" "33316_r_at"
509 "33317_r_at" "33318_r_at"
510 "33319_r_at" "33320_r_at"
511 "33321_r_at" "33322_r_at"
512 "33323_r_at" "33324_r_at"
513 "33325_r_at" "33326_r_at"
514 "33327_r_at" "33328_r_at"
515 "33329_r_at" "33330_r_at"
516 "33331_r_at" "33332_r_at"
517 "33333_r_at" "33334_r_at"
518 "33335_r_at" "33336_r_at"
519 "33337_r_at" "33338_r_at"
520 "33339_r_at" "33340_r_at"
521 "33341_r_at" "33342_r_at"
522 "33343_r_at" "33344_r_at"
523 "33345_r_at" "33346_r_at"
524 "33347_r_at" "33348_r_at"
525 "33349_r_at" "33350_r_at"
526 "33351_r_at" "33352_r_at"
527 "33353_r_at" "33354_r_at"
528 "33355_r_at" "33356_r_at"
529 "33357_r_at" "33358_r_at"
530 "33359_r_at" "33360_r_at"
531 "33361_r_at" "33362_r_at"
532 "33363_r_at" "33364_r_at"
533 "33365_r_at" "33366_r_at"
534 "33367_r_at" "33368_r_at"
535 "33369_r_at" "33370_r_at"
536 "33371_r_at" "33372_r_at"
537 "33373_r_at" "33374_r_at"
538 "33375_r_at" "33376_r_at"
539 "33377_r_at" "33378_r_at"
540 "33379_r_at" "33380_r_at"
541 "33381_r_at" "33382_r_at"
542 "33383_r_at" "33384_r_at"
543 "33385_r_at" "33386_r_at"
544 "33387_r_at" "33388_r_at"
545 "33389_r_at" "33390_r_at"
546 "33391_r_at" "33392_r_at"
547 "33393_r_at" "33394_r_at"
548 "33395_r_at" "33396_r_at"
549 "33397_r_at" "33398_r_at"
550 "33399_r_at" "33400_r_at"
551 "33401_r_at" "33402_r_at"
552 "33403_r_at" "33404_r_at"
553 "33405_r_at" "33406_r_at"
554 "33407_r_at" "33408_r_at"
555 "33409_r_at" "33410_r_at"
556 "33411_r_at" "33412_r_at"
557 "33413_r_at" "33414_r_at"
558 "33415_r_at" "33416_r_at"
559 "33417_r_at" "33418_r_at"
56
```

2006: The Stage is Set



Potti et al (2006, Aug), NEJM, 355:570-80.

Genomic signatures to guide the use of
chemotherapeutics

Anil Potti¹⁻⁷, Holly K Dressman^{1,2}, Andrea Bild^{1,3}, Richard F Riedel^{1,2}, Gina Chan⁴, Robyn Soyer⁴, Daniel Cragan⁵, Hope Cottrell⁶, Michael J Kelley², Rebecca Petersen⁷, David Harpole³, Jeffrey Marks³, Andrew Berchuck^{1,8}, Geoffrey S Ginsburg^{1,2}, Phillip Febbo¹⁻³, Johnathan Lancaster⁴ & Joseph R Nevins¹⁻³

Potti et al (2006, Nov), Nature Medicine, 12:1294-1300.

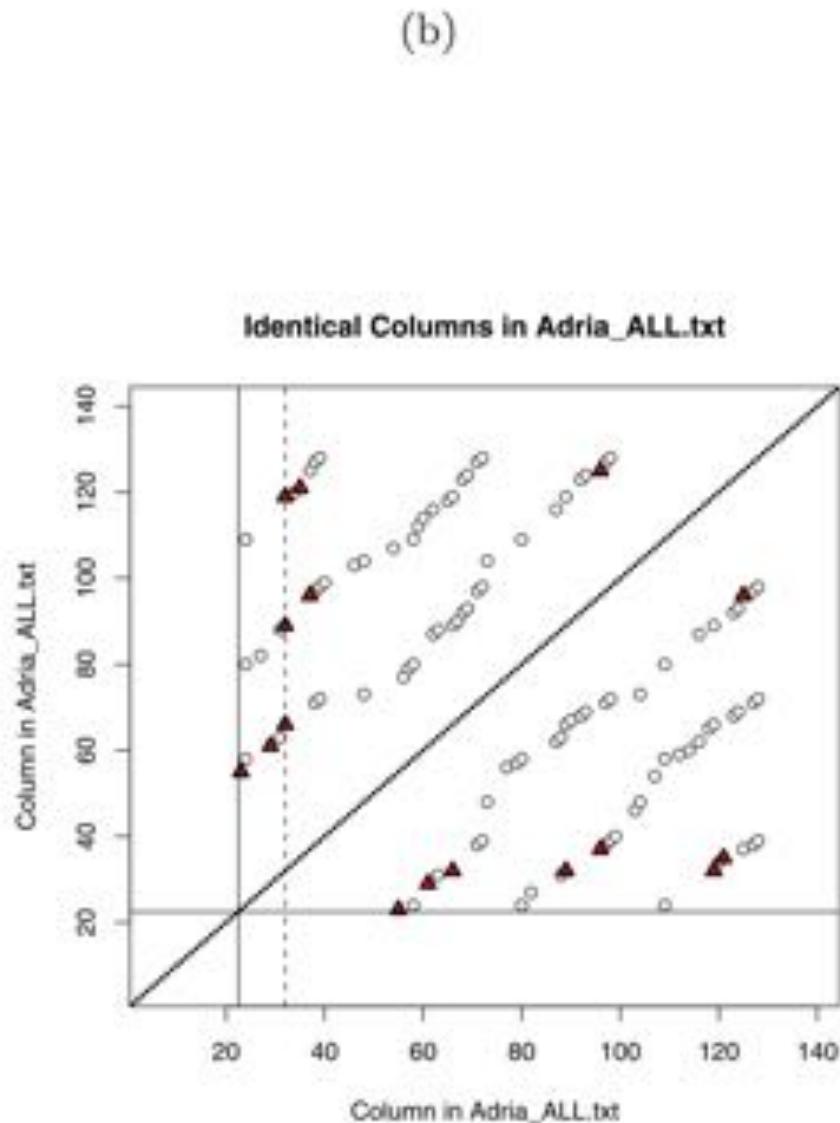
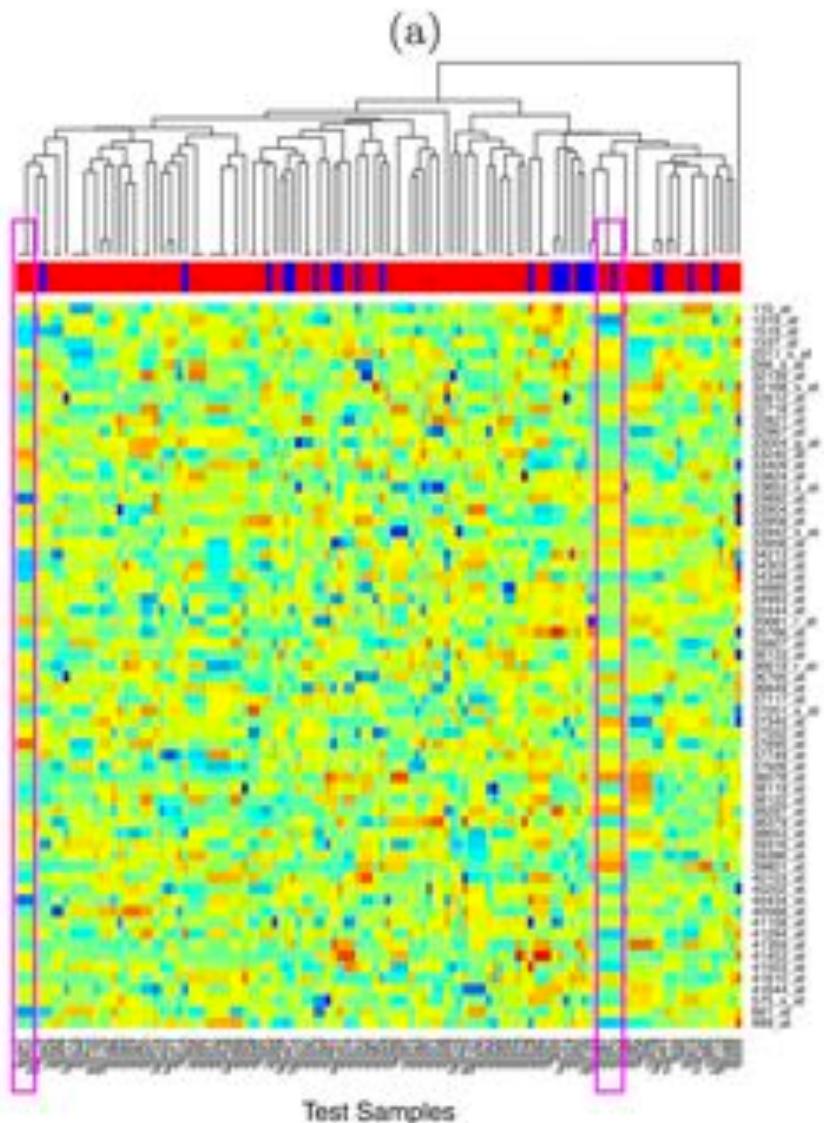
We begin communicating with Potti and Nevins.

Off by One

```
> temp <- cbind(  
+   sort(rownames(pottiUpdated)[fuRows]),  
+   sort(rownames(pottiUpdated)[  
+     fuTQNorm@p.values <= fuCut]));  
> colnames(temp) <- c("Theirs", "Ours");  
> temp
```

	Theirs	Ours
...		
[3,]	"1881_at"	"1882_g_at"
[4,]	"31321_at"	"31322_at"
[5,]	"31725_s_at"	"31726_at"
...		

Dec 4



Clinical Trials

Four clinical trials using the Potti et al Nat Med approach to choose patient therapy were started in 2007-8:

- 3 at Duke
- 1 at Moffitt

A fifth (larger) cooperative group trial (CALGB 30702) in lung cancer was proposed in 2009.

At the same time, a large cooperative group trial (CALGB 30506) testing the Lung Metagene Score (LMS) opened.
The LMS was not guiding therapy.



<http://videolectures.net>

Lecture popularity: ★★★★☆

You need to login to cast your vote.

[Tweet](#) 54 [Like](#) 143 [+1](#) 31 [Share](#) 11 [Print](#) 1

See Also:

The Importance of Reproducibility in High-Throughput Biology: Case Studies in Forensic Bioinformatics

Keith A. Baggerly

Bioinformatics and Computational Biology
UT M. D. Anderson Cancer Center
kabagg@mdanderson.org

Cambridge, 4 September 2010



50dm

- 0:09 The Importance of Reproducibility in High-Throughput Biology: Case Studies in Forensic Bioinformatics
- 1:19 Why is RR So Important in H-TB?

http://videolectures.net/cancerbioinformatics2010_baggerly_irrh/

It's not all about fraud!

You don't need to be
a **statistics wizard** to
spot the problems

It's not all about
the first and last authors

Why should I work reproducibly?

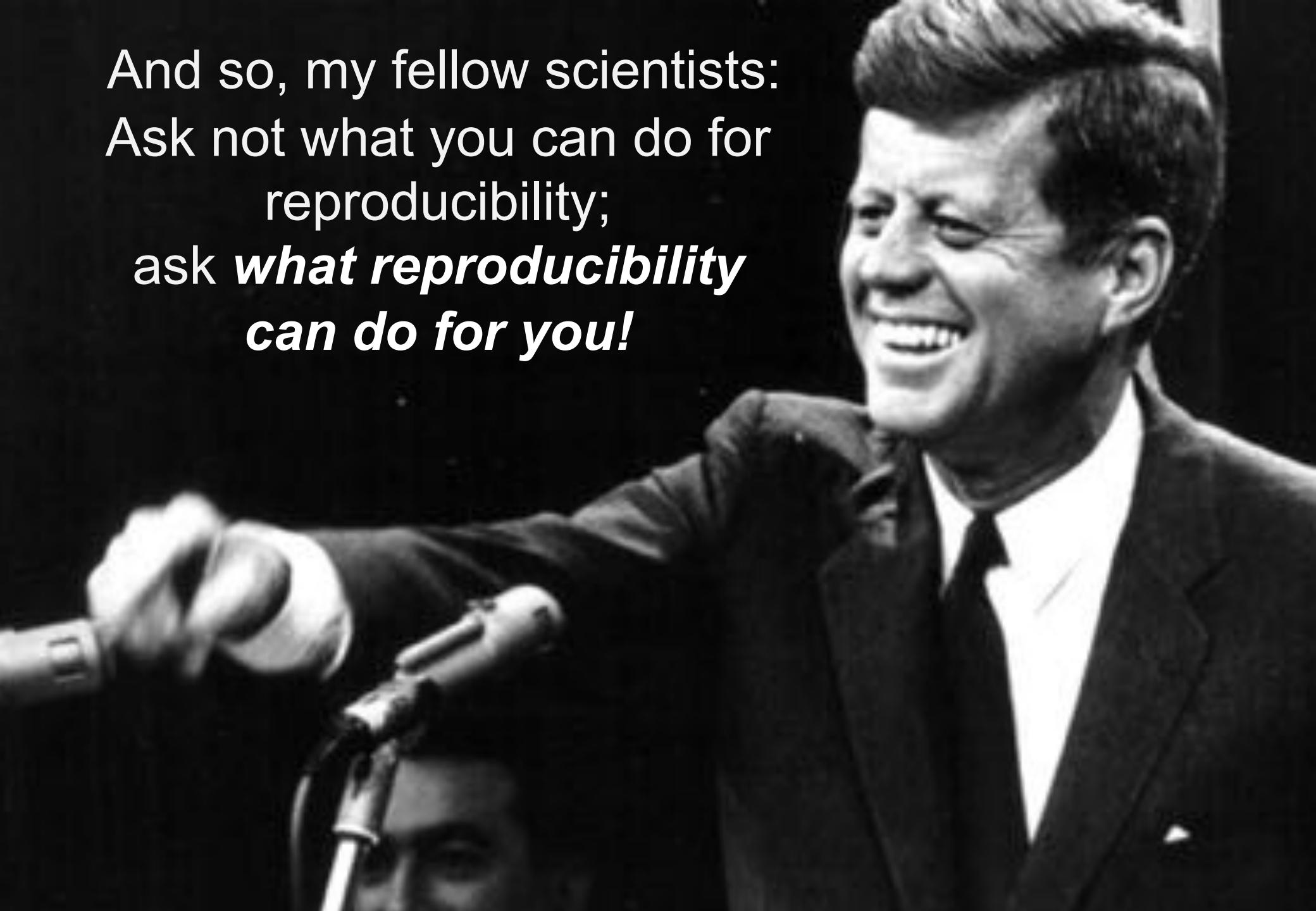
- It's the **right** thing to do!
- It's the **foundation** of Science!
- It's the **honourable** thing to do!
- It's making the world a **better place**!







And so, my fellow scientists:
Ask not what you can do for
reproducibility;
***ask what reproducibility
can do for you!***



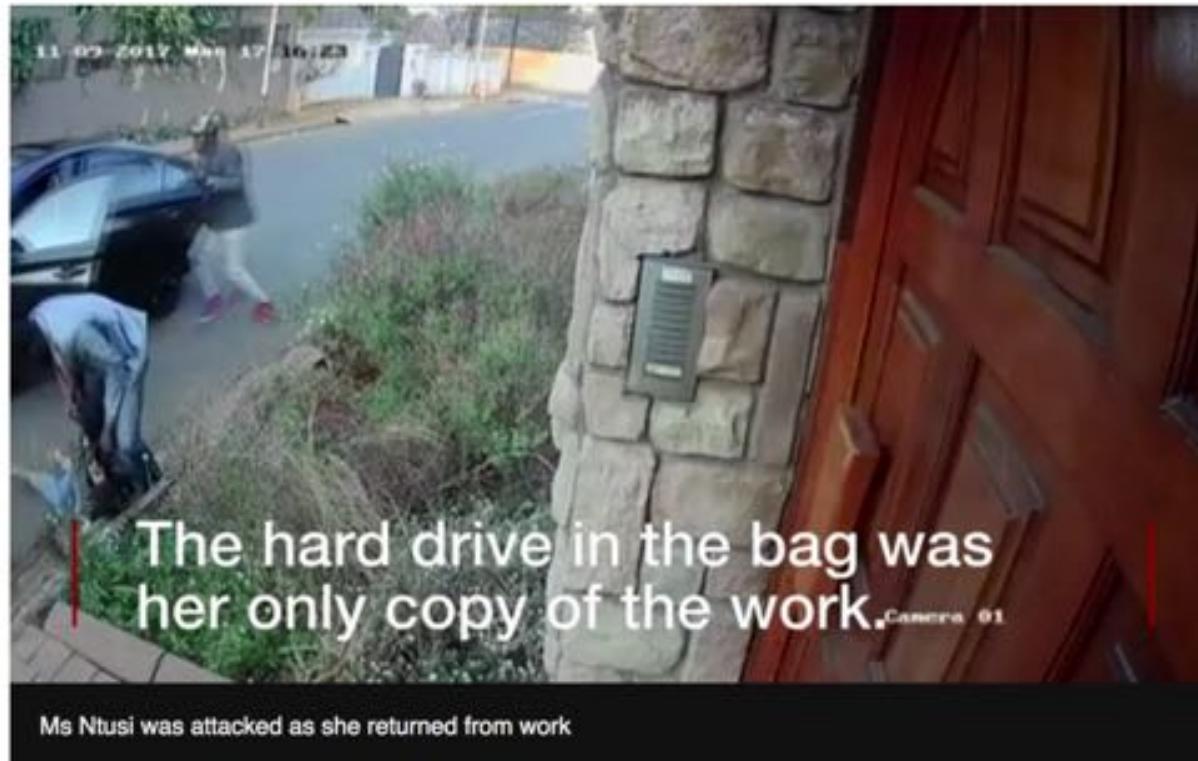


Reproducibility
helps to
avoid disaster

South Africa student fights to keep thesis during robbery

13 September 2017 | [Africa](#)

Share



Very important! LAPTOP LOST in the bus 345



If you find this bag, please return it to the owner, Room 1425, at South Norwegian College,
Route 107 in Fredrikstad, writing a black flag with containing the following:

CRUCIAL scientific data
+ many **YEARS** of
research work inside!

This is sent by Imperial College. Please return and forward
without delay. CALL on me and have no fear!

CASH REWARD

for returning my lost backpack



203 Pavement Lane.com

- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is **VERY** important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on: 07804430054 (ar456@cam.ac.uk)

Thank you!!

SHARE

LETTERS

Editorial expression of concern



0

Jeremy Berg

+ Author Affiliations



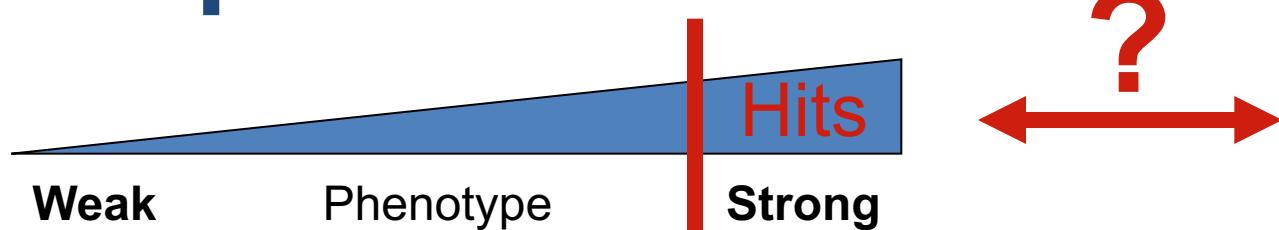
Science 09 Dec 2016:

In the 3 June issue, Science published the Report "Environmentally relevant concentrations of microplastic particles influence larval fish ecology" by Oona M. Lönnstedt and Peter Eklöv (1). The authors have notified Science of the theft of the computer on which the raw data for the paper were stored. These data were not backed up on any other device nor deposited in an appropriate repository. Science is publishing this Editorial Expression of Concern to alert our readers to the fact that no further data can be made available, beyond those already presented in the paper and its supplement, to enable readers to understand, assess, reproduce, or extend the conclusions of the paper.

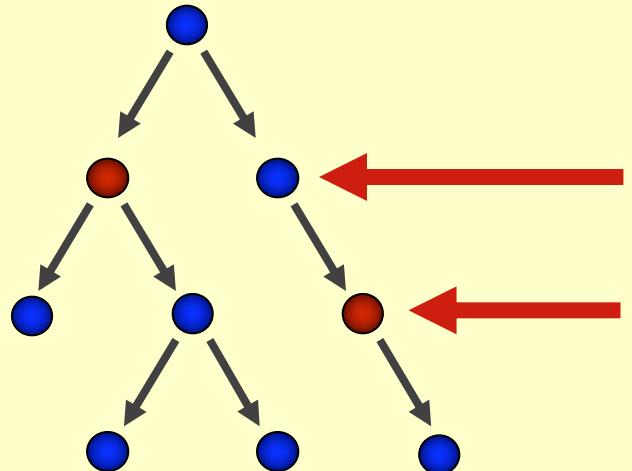
Reference

Anatomy of the NF κ B pathway

Step 1



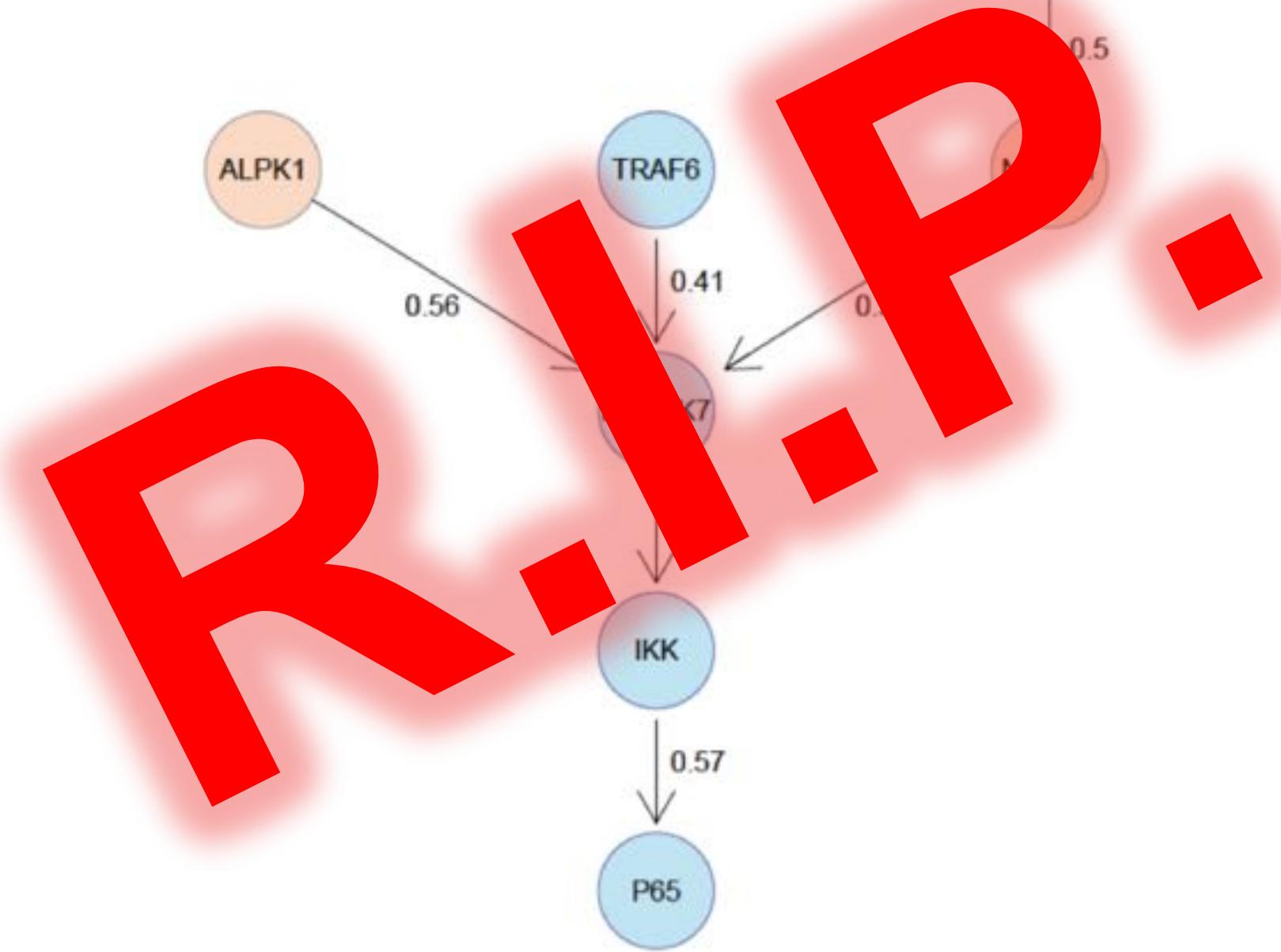
Step 2



**Knock-down
Known pathway
members
New RNAi Hits**

} Compare expression phenotypes by **NEMs**

What a beautiful result!



A publication is more
than a beautiful result!

Starting with
reproducibility early
helps saving time later



**Reproducibility
helps
writing papers**

OPEN ACCESS PEER REVIEWED

RESEARCH ARTICLE

1

2

Spatial and Temporal Heterogeneity in High-Grade Ovarian Cancer: A Phylogenetic Analysis

Discussion

Supporting Information

Acknowledgments

Author Contributions

References

Reader Comments (0)

Media Coverage (1)

Figures

(PDF)

S1 Protocol. Sweave file for survival analysis.

doi:10.1371/journal.pmed.1001789.s020

(RNW)

S1 Table. Distribution of samples.

doi:10.1371/journal.pmed.1001789.s021

(PDF)

S2 Table. Digital PCR results.

Supporting Information

Acknowledgments

Author Contributions

References

High-grade ovarian cancer (HGSC) is a heterogeneous disease that is often resistant to platinum-based chemotherapy. The objective of this study was to determine whether intra-tumour genetic heterogeneity resulting from clonal evolution and the emergence of subclonal tumour populations in HGSC was associated with the development of resistant disease.

Methods and Findings

Evolutionary inference and phylogenetic quantification of heterogeneity was performed using the MEVOCC algorithm on both overall whole genome copy number profiles and selected

PLOS Medicine Cancer Research

Subject Areas

Genome evolution

Phylogenetic analysis

```

148
149 First we load the data and R packages. The data file is part of the paper
150 also made a copy available online.
150 - <<message=FALSE,tidy=FALSE>>-
151 library(survival)
152 library(kernlab)
153 library(rms)
154 library(spatstat)
155 library(RColorBrewer)
156 library(gplots)
157
158 ## load data file from local copy or from URL
159 - if (file.exists("Schwarz2015-supplement.Rdata")){
160   load("Schwarz2015-supplement.Rdata")
161   cat("Data loaded from local copy")
162 - } else {
163   load(url("http://www.markowetzlab.org/supplements/Schwarz2015-supplement.Rdata"))
164   cat("Data loaded from URL") }
165
166
167 The first object in the .Rdata file is a table \texttt{D} with patient
168 - <<message=FALSE>>-
169 D
170 attach(D)
171
172
173 - <<echo=FALSE,message=FALSE>>-
174 ## Print the data table in LaTeX format for inclusion into main manuscript
175 library(xtable)
176 print(xtable(D),file="TableOverview.tex")
177
178
179 Rownames correspond to sample identifiers. Columns indicate the patient
180 identifier (\texttt{Nr}), as well as values for temporal heterogeneity (\texttt{CE}),
181 index (\texttt{CE}), overall survival in days (\texttt{OS}), progressive
182 survival (\texttt{PFS}) and indicators for survival (\texttt{dead}) and progress
183 covariates for age, stage (ordered factor), residual disease after deb
184 1, ordered factor) and the number of samples per patient (\texttt{N}).
185
186

```

1 Clinical data

1.1 Data overview

First we load the data and R packages. The data file is part of the paper supplement, and we have also made a copy available online.

```

library(survival)
library(kernlab)
library(rms)
library(spatstat)
library(RColorBrewer)
library(gplots)

## load data file from local copy or from URL
if (file.exists("Schwarz2015-supplement.Rdata")){
  load("Schwarz2015-supplement.Rdata")
  cat("Data loaded from local copy")
} else {
  load(url("http://www.markowetzlab.org/supplements/Schwarz2015-supplement.Rdata"))
  cat("Data loaded from URL") }

## Data loaded from URL.

```

The first object in the .Rdata file is a table D with patient information:

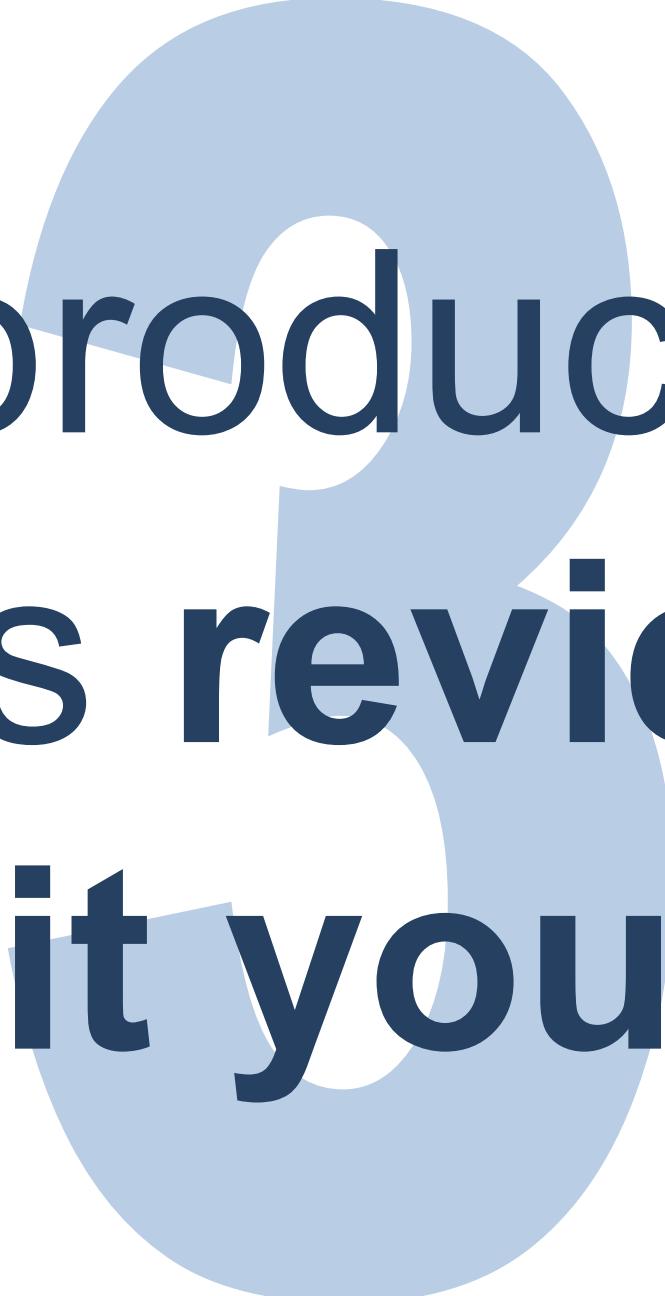
D

	Nr	TH	CE	OS	PFS	dead	prog	Mist	Age	Stage	residual
##	OV03-01	1	4.730231	1.2865274	511	271	1	1	62	IV	<1cm
##	OV03-02	2	NA	0.7106901	977	363	1	1	62	IV	<1cm
##	OV03-04	3	3.735366	1.2432629	209	153	1	1	69	IV	>1cm
##	OV03-07	4	NA	NA	626	616	1	1	62	IIIC	Nil
##	OV03-08	5	3.801712	1.4706631	547	303	1	1	63	IV	<1cm
##	OV03-10	6	6.588895	0.7298828	744	298	1	1	62	IV	<1cm
##	OV03-13	7	3.000290	0.6836961	1587	358	1	1	61	IV	>1cm
##	OV03-17	8	3.423112	2.2357817	889	373	1	1	61	IIIC	<1cm
##	OV03-20	9	4.487828	0.6494353	1278	563	1	1	71	IV	>1cm
##	OV03-21	10	4.719848	0.6664309	1139	303	1	1	60	IIIC	>1cm
##	OV03-22	11	5.702720	0.4834086	1856	382	1	1	58	IIIC	<1cm
##	OV03-23	12	NA	NA	1565	634	1	1	60	IIIC	Nil
##	OV03-24	NA	NA	NA	376	375	1	1	63	IIIC	>1cm
##	OV03-25	13	NA	0.6215297	1166	776	1	1	57	IIIC	>1cm
##	OV04-20	14	4.621984	0.6083119	1513	601	0	1	63	IIIC	Nil
##	OV04-21	15	NA	0.7412773	706	332	1	1	54	IV	Nil
##	OV04-27	16	NA	NA	1408	1408	0	0	58	IIIC	Nil
##	OV04-30	17	NA	0.8591205	849	293	1	1	60	IIIC	>1cm
##											

Why is well-documented and easily accessible code+data useful?

- Easy to look up numbers and put them in paper
- Results automatically update when data change.
- It is engaging: more eye contact with the data analysis.
- Easier to spot mistakes.

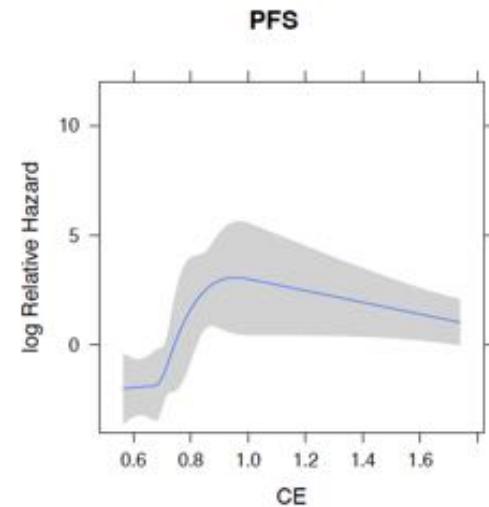
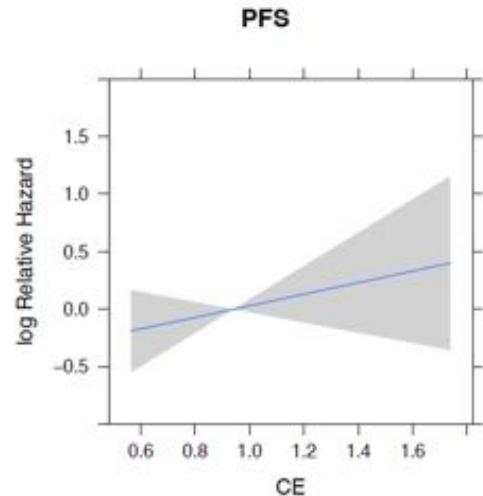
```
> table(stage_expected)
stage_expected
 1 2 3 4
17 10 5 10
> table(stage_observed)
stage_observed
 2 3 4 999 John XXX
20 10 5 2 1 3
```



Reproducibility
helps reviewers
see it your way

A very engaged reviewer

- **Reviewer:** “I downloaded the authors’ data and tried out a variation of their analysis which gave an insignificant result”



- **We:** “Thank you, the reason is XXX and if you do YYY everything is fine.”



**Reproducibility
enables
continuity**

*“My boss said I should
continue the project of a
previous postdoc.*

*But that postdoc is long gone
and left no scripts or data.”*

**“I’m sorry,
*I did this analysis
6 months ago.”***

*“I am so busy,
I can’t remember all
the details of all my
projects”*

The first re-user of your data
will be your **future-self**.

And **past-self** does
not answer emails.

Reproducibility
helps to build
your
reputation

Search:

[Home](#) > [Bioconductor 3.1](#) > [Experiment Packages](#) > Fletcher2013a

Fletcher2013a

platforms [all](#) | downloads [available](#) | posts [0](#)
build [ok](#) | commits [0.83](#)

Gene expression data from breast cancer cells under FGFR2 signalling perturbation.

Bioconductor version: Release (3.1)

The package `Fletcher2013a` contains time-course gene expression data from MCF-7 cells treated under different experimental systems in order to perturb FGFR2 signalling. The data comes from Fletcher et al. (*Nature Comms* 4:2464, 2013) where further details about the background and the experimental design of the study can be found.

Author: Mauro Castro, Michael Fletcher, Florian Markowetz and Kerstin Meyer.

Maintainer: Mauro Castro <mauro.a.castro@gmail.com>

Citation (from within R, enter `citation("Fletcher2013a")`):

Fletcher M, Castro M, Wang X, Santiago Id, O'Reilly M and al. e (2013). "Master regulators of FGFR2 signalling and breast cancer risk." *Nature Communications*, 4, pp. 2464.

[Installation](#)

Workflows >

Common Bioconductor workflows include:

- [Oligonucleotide Arrays](#)
- [High-throughput Sequencing](#)
- [Counting Reads for Differential Expression](#) (`parathyroideSE` vignette)
- [Annotation](#)
- [Annotating Variants](#)
- [Annotating Ranges](#)
- [Flow Cytometry and other assays](#)
- [Candidate Binding Sites for Known Transcription Factors](#)
- [Cloud-enabled cis-eQTL search and annotation](#)
- [RNA-Seq workflow: gene-level exploratory analysis and differential expression](#)
- [Changing genomic coordinate systems with `rtracklayer::liftOver`](#)
- [Mass spectrometry and proteomics data analysis](#)

Mailing Lists >

Post questions about Bioconductor packages to our mailing lists. Read the [posting guide](#) before posting!

Promoting an open research culture

Author guidelines for journals could help to promote transparency, openness, and reproducibility

By B. A. Nosek,¹ G. Alter,² G. C. Banks,³ D. Borsboom,⁴ S. D. Bowman,⁵ S. J. Breckler,⁶ S. Buck,⁷ C. D. Chambers,⁸ G. Chin,⁹ G. Christensen,¹⁰ M. Contestabile,¹¹ A. Dafoe,¹² E. Eich,¹³ J. Freese,¹⁴ R. Glennerster,¹⁵ D. Goroff,¹⁶ D. P. Green,¹⁷ B. Hesse,¹⁸ M. Humphreys,¹⁹ J. Ishiyama,²⁰ D. Karlan,²¹ A. Kraut,²² A. Lupia,²³ P. Mabry,²⁴ T. A. Madon,²⁵ N. Malhotra,²⁶ E. Mayo-Wilson,²⁷ M. McNutt,²⁸ E. Miguel,²⁹ E. Levy Paluck,³⁰ U. Simonsohn,³¹ C. Soderberg,³² B. A. Spellman,³³ J. Turitto,³⁴ G. VandenBos,³⁵ S. Vazire,³⁶ E. J. Wagenmakers,³⁷ R. Wilson,³⁸ T. Yarkoni³⁹

Transparency, openness, and reproducibility are readily recognized as vital features of science (*1, 2*). When asked, most scientists embrace these features as disciplinary norms and values (*3*). Therefore, one might expect that these valued features would be routine in daily practice. Yet, a growing body of evidence suggests that this is not the case (*4–6*).

Downloaded from www.sciencemag.org on July 3, 2015

Summary of the eight standards and three levels of the TOP guidelines

Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Research materials transparency	Journal encourages materials sharing—or says nothing.	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Preregistration of studies	Journal says nothing.	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing.	Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists.	Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies—or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts blind review of results.	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.

5 selfish reasons to work reproducibly

1. Avoid disaster
2. Easier to write papers
3. Easier to talk to reviewers
4. Continuity of your work
5. Reputation



So What?

 You and 82 others don't give a fuck.

**“Just because you
do open science
you are not
necessarily right”**

**“All this reproducibility
stuff is just for
nerds with OCD”**

“It’s only the
result that
matters!”

“I’d rather do
real science
than tidy up
my data”

**“We can always sort
out the code and data
after submission”**

**“My field is very
competitive
and I can’t risk
wasting time”**

**“If there is
wetlab validation,
who cares if the
computer stuff is
reproducible?”**

“Excel works just fine.

**I don’t need any fancy R
or Python or whatever.”**

**“Mind your own
business!**

**I document my data
the way I want!”**

“The current
reward structure
does not enough to
incentivize
data sharing”

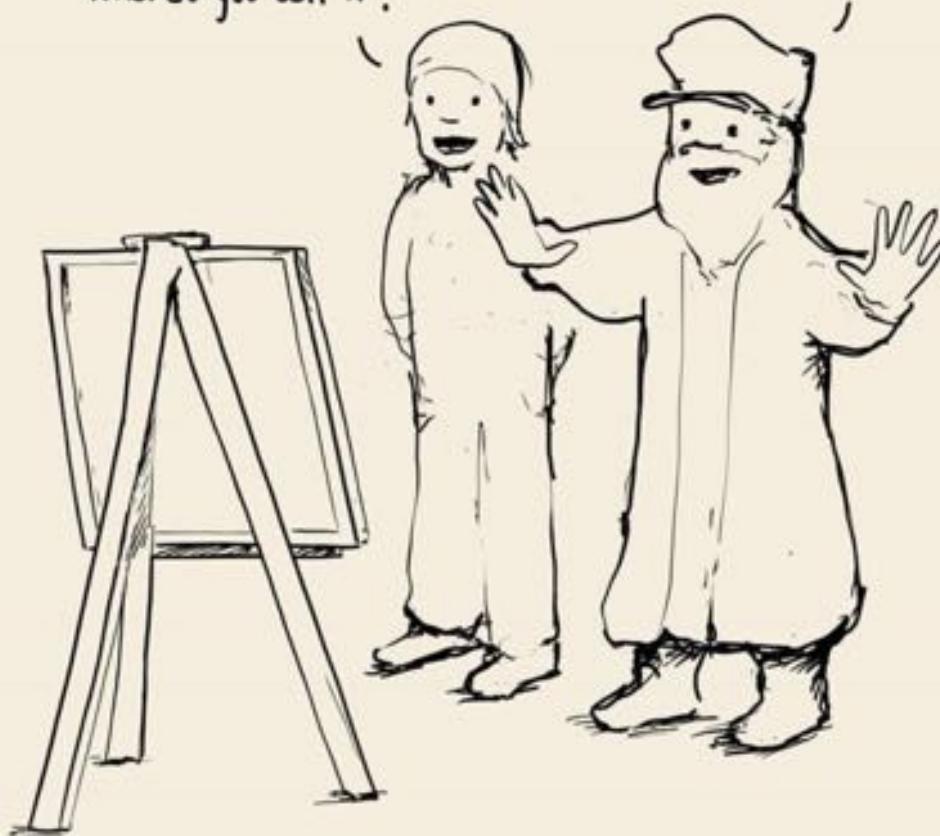
A young child with blonde hair, wearing a blue t-shirt and dark pants, is sitting on a stack of approximately ten books. The books are of various colors, including white, black, and grey. The child is looking down at the books. The background is a plain, light-colored wall.

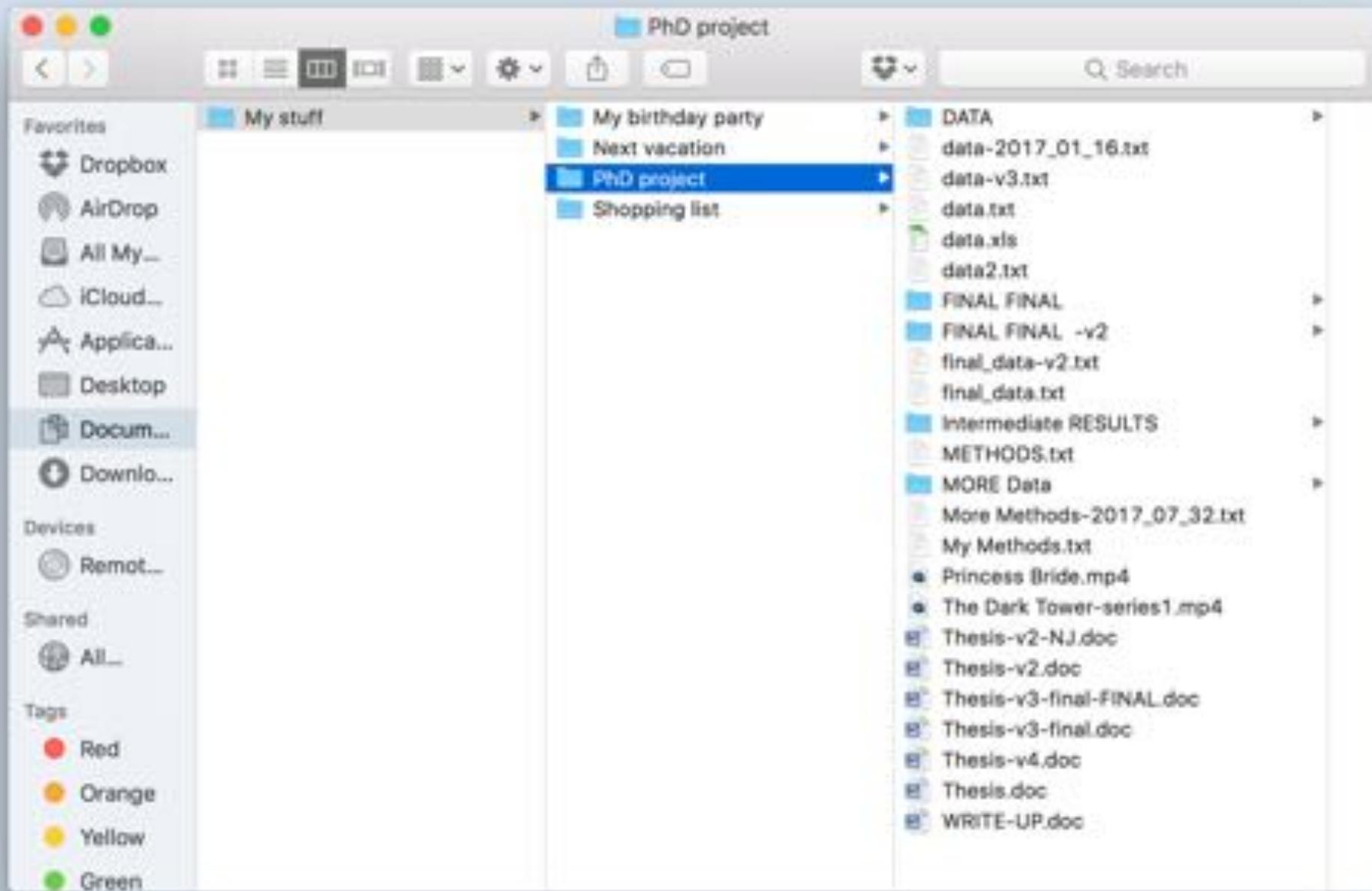
How to get started?

Baby steps towards
reproducibility

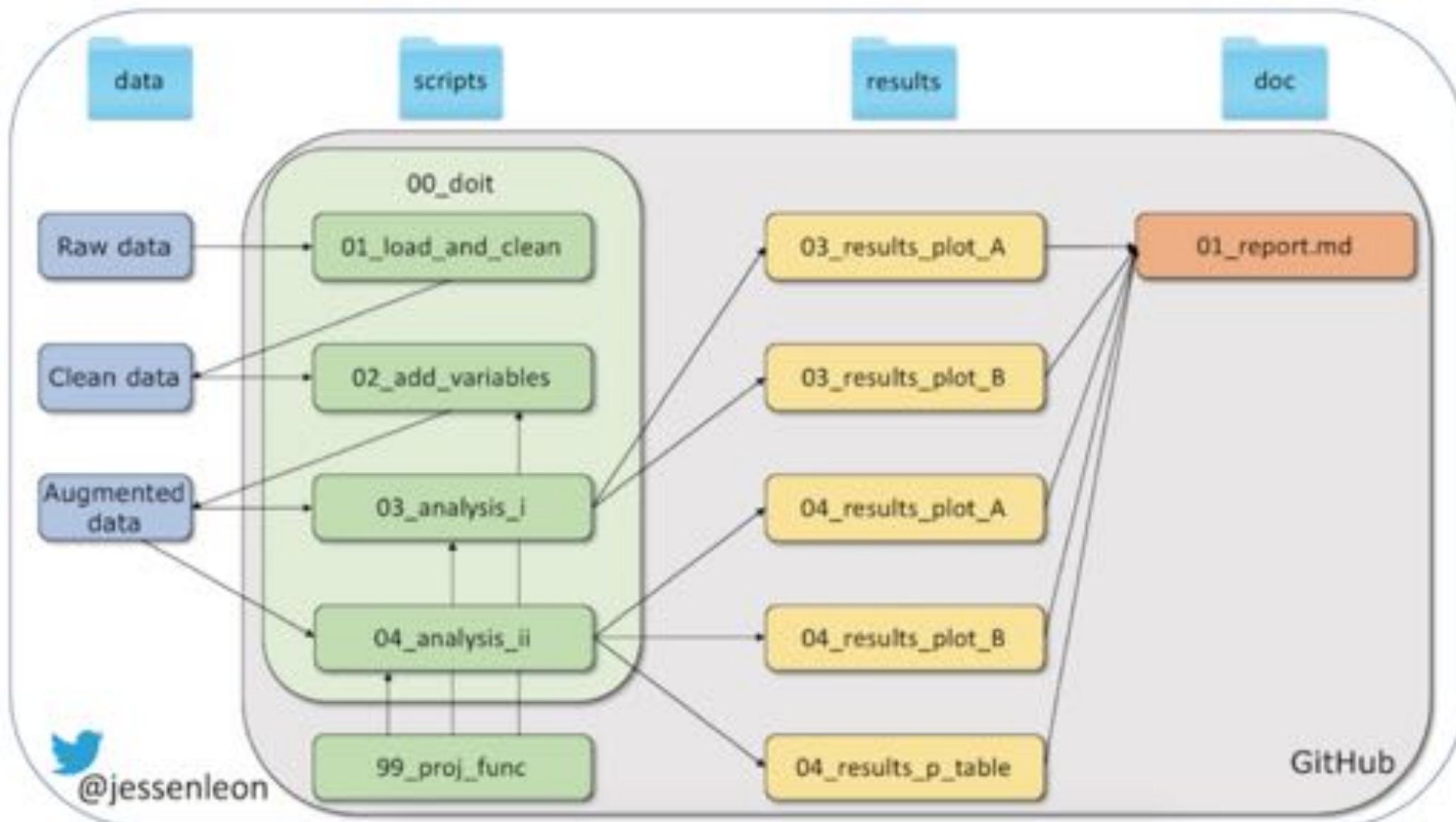
beautiful!
what do you call it?

mona_lisa_finalrealupdateFINALL6





Tidy folder structure



A

Untidy Data

species	habitat	weight	length	latitude/longitude	date
Alligator mississippiensis	swamp	431 lb	4 ft 2	29.531,-82.184	Sept 15, 2015
Puma concolor	forest	125 lb	2.2m	29.125,-81.682	08/10/2015
Ursus americanus	forest	88 kg	133 cm	N29°7'30"/W81°40'55.2"	07-13-2015

B

Tidy Data

meta-data

data

species_code	date	station_code	weight_kg	length_cm
TSN 551771	2015-09-15	1	196	127
TSN 55247	2015-08-10	2	57	220
TSN 180544	2015-07-13	2	88	133

station_code	habitat	latitude	longitude
1	swamp	29.531	-82.184
2	forest	29.125	-81.682

species_code	class	genus	species
TSN 551771	Reptilia	Alligator	mississippiensis
TSN 55247	Mammalia	Puma	concolor
TSN 180544	Mammalia	Ursus	americanus

When a paper is accepted in a top-tier journal

“Any Supplementary Tables to be published on their own as Excel or CSV files in Nature Genetics should ideally be **modeled on Tidy Data standards**:

<https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html>”

Less clicking and pasting,
more scripting and coding

So why is there
still resistance?





eLIFE

ABOUT LABS COMMUNITY

SUBMIT MY RESEARCH

LOG IN/REGISTER

≡ MENU

HOME

MAGAZINE



CURATED BY Roger Davis et al.

Reproducibility Project: Cancer Biology



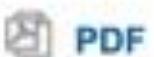
COLLECTION Dec 10, 2014

The Reproducibility Project: Cancer Biology is an initiative to independently replicate selected results from a number of high-profile papers in the field of cancer biology. For each paper a Registered Report detailing the proposed experimental designs and protocols for the replications is peer reviewed and published prior to data collection; the results of these experiments are then published as a Replication Study. The project is a collaboration between the Center for Open Science and Science Exchange.

If a researcher spends six months, say, trying to replicate [highly specialised] work and reports that it is **irreproducible**, that

- **can deter other scientists** from pursuing a promising line of research,
- jeopardize the original scientists' **chances of obtaining funding** to continue it themselves,
- and potentially **damage their reputations**.

reputations of careful, meticulous scientists.



PDF



Rights & Permissions

Subject terms: Cell biology - Publishing

Archive | Audio & Video | For Au



lication drive

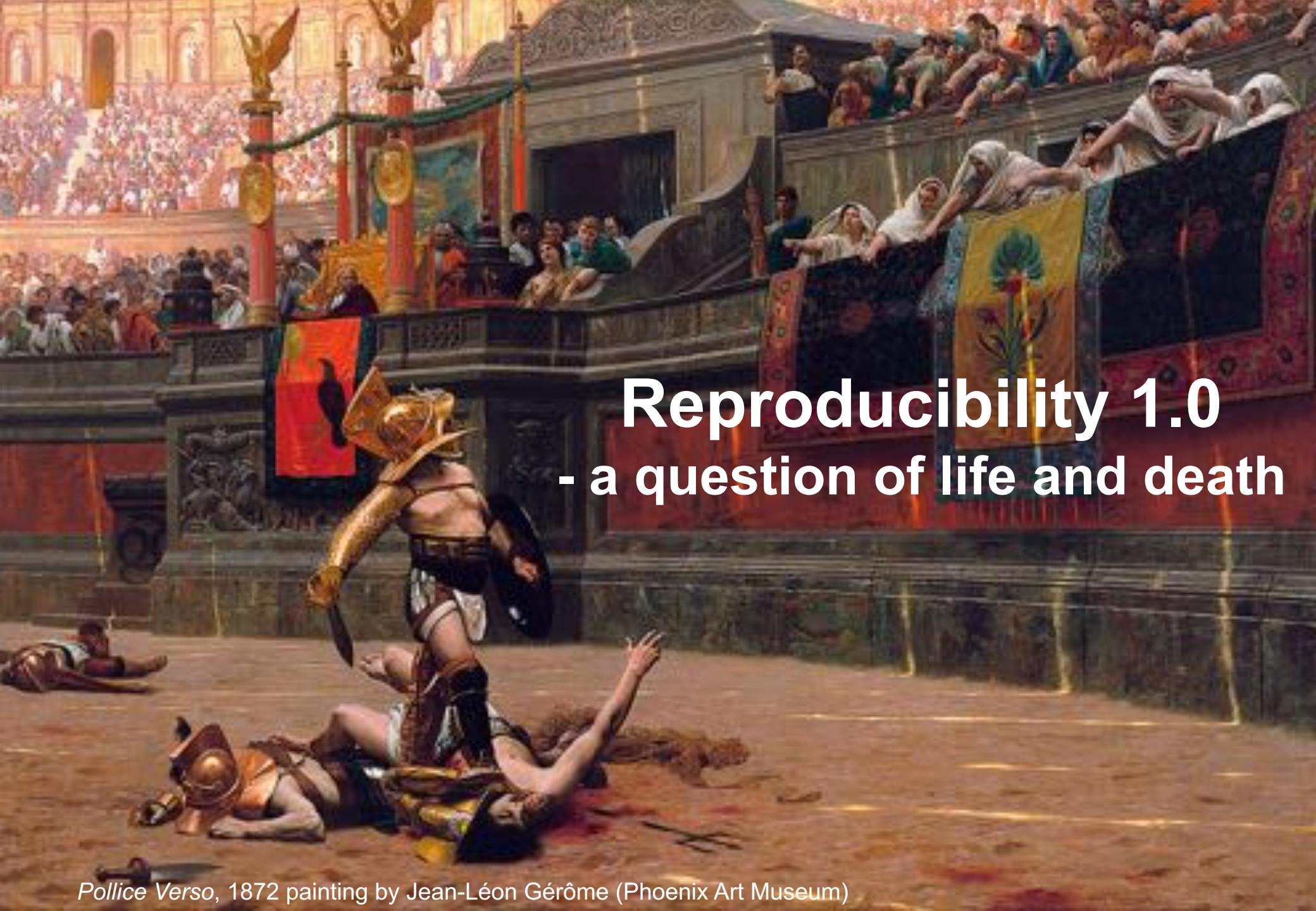
[A]ll journals should set aside a small space to publish short, peer-reviewed reports from groups that **get together to collaboratively solve reproducibility problems**, describing their trials and tribulations in detail.



JUDGE DREDD



REPLICATION POLICE

The background is a painting titled "Pollice Verso" by Jean-Léon Gérôme, depicting a gladiatorial combat scene in an arena. In the foreground, a gladiator in a blue and white tunic is fallen on the ground, holding a shield with a red emblem featuring a black bird. Another gladiator in a red and gold tunic is kneeling over him, having just delivered a blow with a sword. In the background, spectators in the stands watch the fight. A large golden eagle statue sits atop a column in the upper left. The text "Reproducibility 1.0 - a question of life and death" is overlaid on the right side of the painting.

Reproducibility 1.0

- a question of life and death



Improving Openness & Reproducibility of Scientific Research

Dr Timothy Errington

Improving Openness and Reproducibility of Scientific Research

Tim Errington
Center for Open Science
<http://cos.io/>



Reproducibility 2.0

- overcoming obstacles



Sorting Out the FACS: A Devil in the Details

William C. Hines,^{1,5,*} Ying Su,^{2,3,4,5,*} Irene Kuhn,¹ Kornelia Polyak,^{2,3,4,5} and Mina J. Bissell^{1,5}

¹Life Sciences Division, Lawrence Berkeley National Laboratory, Mailstop 977R225A, 1 Cyclotron Road, Berkeley, CA 94720, USA

²Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

³Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

⁴Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

⁵These authors contributed equally to this work

*Correspondence: chines@lbl.gov (W.C.H.), ying_su@dfci.harvard.edu (Y.S.)

<http://dx.doi.org/10.1016/j.celrep.2014.02.021>

The reproduction of results is the cornerstone of science; yet, at times, reproducing the results of others can be a difficult challenge. Our two laboratories, one on the East and the other on the West Coast of the United States, decided to collaborate on a problem of mutual interest—namely, the heterogeneity of the human breast. Despite using seemingly identical methods, reagents, and specimens, our two laboratories quite reproducibly were unable to replicate each other's fluorescence-activated cell sorting (FACS) profiles of primary breast cells. Frustration mounted, given that we had not found the correct answer(s), even after a year. Rather than giving up or each publishing

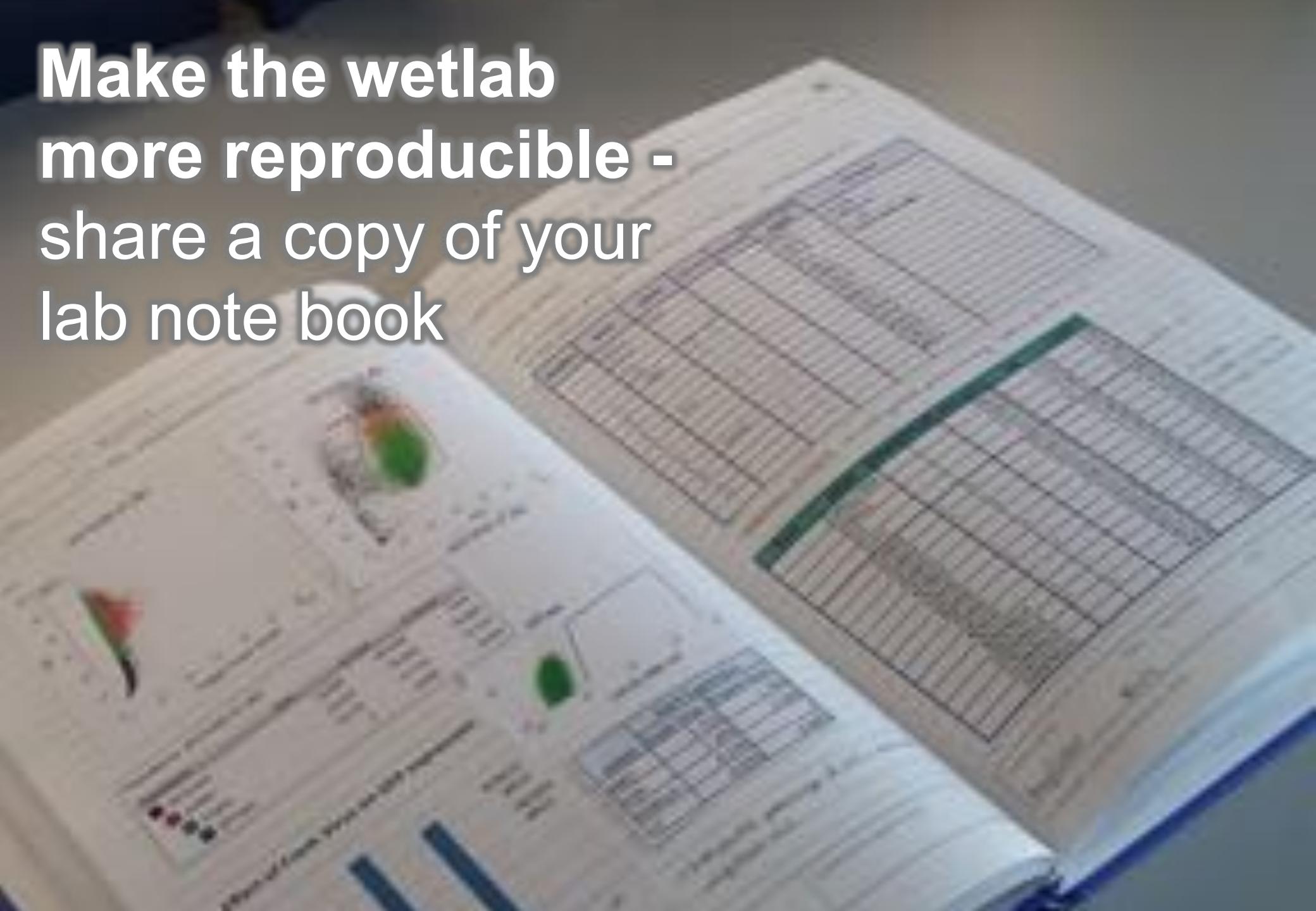
of studying cells close to their context *in vivo* makes the exercise even more challenging.

Paired with *in situ* characterizations, FACS has emerged as the technology most suitable for distinguishing diversity among different cell populations in the mammary gland. Flow instruments have evolved from being able to detect only a few parameters to those now capable of measuring up to—and beyond—an astonishing 50 individual markers per cell (Cheung and Utz, 2011). As with any exponential increase in data complexity, the importance of developing robust preparation and analytical protocols that generate reproducible results increases

breast reduction mammoplasties. Molecular analysis of separated fractions was to be performed in Boston (K.P.'s laboratory, Dana-Farber Cancer Institute, Harvard Medical School), whereas functional analysis of separated cell populations grown in 3D matrices was to take place in Berkeley (M.J.B.'s laboratory, Lawrence Berkeley National Lab, University of California, Berkeley). Both our laboratories have decades of experience and established protocols for isolating cells from primary normal breast tissues as well as the capabilities required for flow sorting primary cells from mice and women.

We settled on isolating cell populations

**Make the wetlab
more reproducible -
share a copy of your
lab note book**



Embrace reproducibility!

It will help you
understand your own
science better

5 selfish reasons to work reproducibly

1. Avoid disaster
2. Easier to write papers
3. Easier to talk to reviewers
4. Continuity of your work/in the lab
5. Reputation
6. You learn some new science

When do you need to worry about reproducibility?

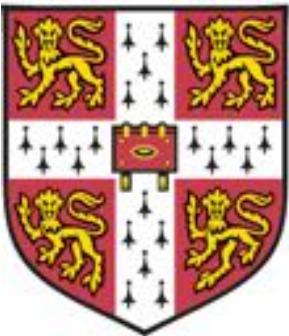
- Before you start the project
 - While you do the analysis
 - When you write the paper
 - When you co-author a paper
 - After you have published a paper
 - When you review a paper
- ALWAYS!**

Who makes reproducibility possible?

- Phd students
 - Postdocs
 - PIs
- Learn tools and apply in daily work!
- Create a '**culture of reproducibility**' in your lab!



5 selfish reasons to work reproducibly



Florian Markowetz
CRUK Cambridge Institute
www.markowetzlab.org
 @markowetzlab



How to transparently improve published papers?





Scientific publications

- The final result
- Static record of scientific achievement
- Hard to update
- Don't look back:
- The last paper is less important than the next one



Genome Biology

[Home](#) [About](#) [Articles](#) [Submission Guidelines](#)**SHARE**[RESEARCH ARTICLE](#)[CANCER](#)

Quantitative ... in Breast Tum

Yinyin Yuan^{1,2,*†}, Henrik Fallm

* See all authors and affiliations

Science Translational Medicine
Vol. 4, Issue 157, pp. 157ra143
DOI: 10.1126/scitranslmed.3004[Article](#)[Figure](#)**This article has a co**[A Correction to the Rese](#)
[Breast Tumors Complex](#)
[S.-F. Chin, R. F. Schwarz,](#)
[Provenzano, S. Aparicio,](#)

Abstract

Solid tumors are heter...
which complicates the...
architecture is general...

[Research](#) | [Open Access](#) | Published: 13 May 2019

VULCAN integrates ChIP-seq with patient-derived co-expression networks to identify GRHL2 as a key co-regulator of ERα at enhancers in breast cancer

Andrew N. Holding , Federico M. Giorgi, Amanda Donnelly, Amy E. Cullen, Sankari Nagarajan, Luke A. Seith & Florian Markowetz

Genome Biology 20, Article number: 91 (2019) | Download Citation 

 The [Correction to this article](#) has been published in Genome Biology 2019 20:122

Abstract

Background

VirtUaL ChIP-seq Analysis through Networks (VULCAN) infers regulatory

[BEST PRODUCTS](#)[REVIEWS](#)[NEWS](#)[VIDEO](#)[HOW TO](#)[SMART HOME](#)[CARS](#)[DEALS](#)[JOIN / SIGN IN](#)[SECURITY](#) | [LEER EN ESPAÑOL](#)

Google will now pay up to \$30,000 for reporting a Chrome bug

You can earn bigger bucks by becoming a digital bounty hunter.

BY SHELBY BROWN, STEPHEN SHANKLAND | JULY 18, 2019 12:11 PM PDT



Ad closed by Google

[Stop seeing this ad](#)

[Why this ad? >](#)

Everyone makes mistakes



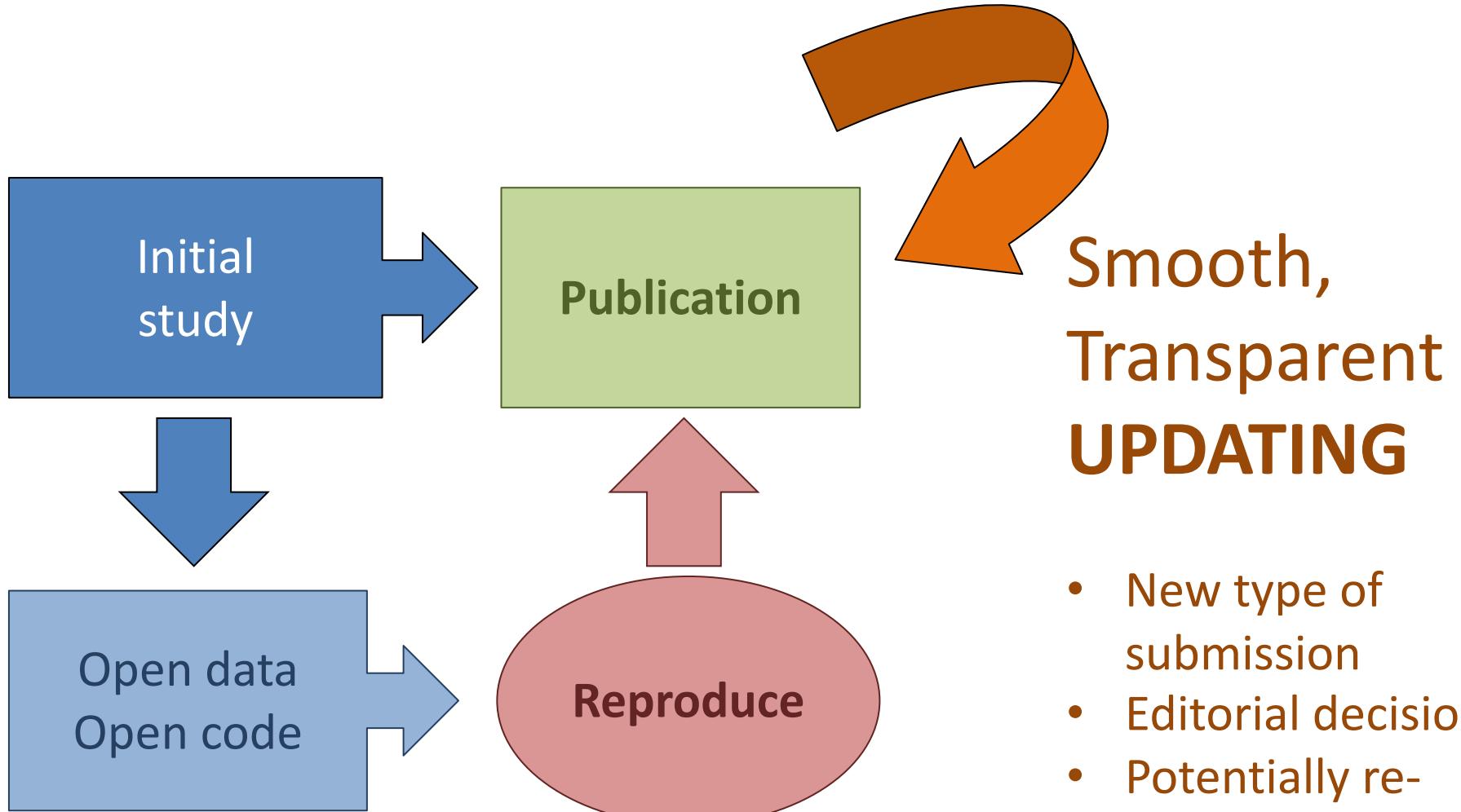
Major

- Expression of concern
- Correction
- Retraction

Minor

Need new ways to smoothly and transparently maintain publications

Reproducibility 3.0



Transparent ‘Bug-fixes’

Publication history

Received: April 3, 2017

Accepted: August 1, 2017

Version of Record published: [September 3, 2017 \(version 1\)](#)

Minor amendment published: [June 17, 2018 \(version 2\)](#)

Minor amendment published: [April 1, 2019 \(version 3\)](#)

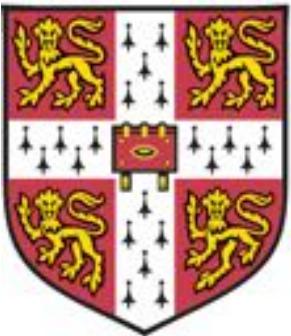


5 selfish reasons to work reproducibly

1. Avoid disaster
2. Easier to write papers
3. Easier to talk to reviewers
4. Continuity of your work/in the lab
5. Reputation
6. You learn some new science
7. You keep your papers bug-free



7 selfish reasons to work reproducibly



Florian Markowetz
CRUK Cambridge Institute
www.markowetzlab.org
 @markowetzlab



Publishing Better Science through Better Data 2016 (#scidata16)



Cartoon by
Royston Robertson
www.roystoncartoons.com/



Courses ▾ Programs ▾ Schools & Partners About ▾

Search:



HOME ▾ ABOUT

ch

Home > All Subjects > Data Analysis & Statistics > Principles, Statistical and Computational Tools for Reproducible Science



Principles, Statistical and Computational Tools for Reproducible Science

Learn skills and tools that support data science and reproducible research, to ensure you can trust your own research results, reproduce them yourself, and communicate them to others.



research is said to be reproducible. Computers play a vital role in many research

A scenic beach scene with lush green trees in the background, a sandy path leading towards the ocean, and clear blue water in the foreground.

Science is a beach

Dry and Wet

RESEARCH MATTERS

All biology is computational biology

Florian Markowetz*

University of Cambridge, Cancer Research UK Cambridge Institute, Cambridge, United Kingdom

* florian.markowetz@cruk.cam.ac.uk

Abstract

Here, I argue that computational thinking and techniques are so central to the quest of understanding life that today all biology is computational biology. Computational biology brings order into our understanding of life, it makes biological concepts rigorous and testable, and it provides a reference map that holds together individual insights. The next modern synthesis in biology will be driven by mathematical, statistical, and computational methods being absorbed into mainstream biological training, turning biology into a quantitative science.

```

148
149 First we load the data and R packages. The data file is part of the paper
150 also made a copy available online.
150 - <<message=FALSE,tidy=FALSE>>-
151 library(survival)
152 library(kernlab)
153 library(rms)
154 library(spatstat)
155 library(RColorBrewer)
156 library(gplots)
157
158 ## load data file from local copy or from URL
159 - if (file.exists("Schwarz2015-supplement.Rdata")){
160   load("Schwarz2015-supplement.Rdata")
161   cat("Data loaded from local copy")
162 - } else {
163   load(url("http://www.markowetzlab.org/supplements/Schwarz2015-supplement.Rdata"))
164   cat("Data loaded from URL") }
165
166
167 The first object in the .Rdata file is a table \texttt{D} with patient
168 - <<message=FALSE>>-
169 D
170 attach(D)
171
172
173 - <<echo=FALSE,message=FALSE>>-
174 ## Print the data table in LaTeX format for inclusion into main manuscript
175 library(xtable)
176 print(xtable(D),file="TableOverview.tex")
177
178
179 Rownames correspond to sample identifiers. Columns indicate the patient
180 identifier (\texttt{Nr}), as well as values for temporal heterogeneity (\texttt{CE}),
181 index (\texttt{CE}), overall survival in days (\texttt{OS}), progressive
182 survival (\texttt{PFS}) and indicators for survival (\texttt{dead}) and progress
183 covariates for age, stage (ordered factor), residual disease after deb
184 1, ordered factor) and the number of samples per patient (\texttt{N}).
185
186

```

1 Clinical data

1.1 Data overview

First we load the data and R packages. The data file is part of the paper supplement, and we have also made a copy available online.

```

library(survival)
library(kernlab)
library(rms)
library(spatstat)
library(RColorBrewer)
library(gplots)

## load data file from local copy or from URL
if (file.exists("Schwarz2015-supplement.Rdata")){
  load("Schwarz2015-supplement.Rdata")
  cat("Data loaded from local copy")
} else {
  load(url("http://www.markowetzlab.org/supplements/Schwarz2015-supplement.Rdata"))
  cat("Data loaded from URL") }

## Data loaded from URL.

```

The first object in the .Rdata file is a table D with patient information:

D

	Nr	TH	CE	OS	PFS	dead	prog	Mist	Age	Stage	residual
##	OV03-01	1	4.730231	1.2865274	511	271	1	1	62	IV	<1cm
##	OV03-02	2	NA	0.7106901	977	363	1	1	62	IV	<1cm
##	OV03-04	3	3.735366	1.2432629	209	153	1	1	69	IV	>1cm
##	OV03-07	4	NA	NA	626	616	1	1	62	IIIC	Nil
##	OV03-08	5	3.801712	1.4706631	547	303	1	1	63	IV	<1cm
##	OV03-10	6	6.588895	0.7298828	744	298	1	1	62	IV	<1cm
##	OV03-13	7	3.000290	0.6836961	1587	358	1	1	61	IV	>1cm
##	OV03-17	8	3.423112	2.2357817	889	373	1	1	61	IIIC	<1cm
##	OV03-20	9	4.487828	0.6494353	1278	563	1	1	71	IV	>1cm
##	OV03-21	10	4.719848	0.6664309	1139	303	1	1	60	IIIC	>1cm
##	OV03-22	11	5.702720	0.4834086	1856	382	1	1	58	IIIC	<1cm
##	OV03-23	12	NA	NA	1565	634	1	1	60	IIIC	Nil
##	OV03-24	NA	NA	NA	376	375	1	1	63	IIIC	>1cm
##	OV03-25	13	NA	0.6215297	1166	776	1	1	57	IIIC	>1cm
##	OV04-20	14	4.621984	0.6083119	1513	601	0	1	63	IIIC	Nil
##	OV04-21	15	NA	0.7412773	706	332	1	1	54	IV	Nil
##	OV04-27	16	NA	NA	1408	1408	0	0	58	IIIC	Nil
##	OV04-30	17	NA	0.8591205	849	293	1	1	60	IIIC	>1cm
##											

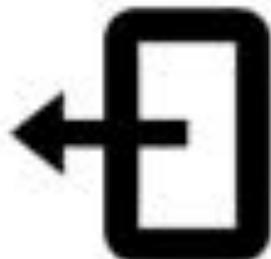
In case of fire



1. git commit



2. git push



3. leave building

Artificial intelligence faces reproducibility crisis

Matthew Hutson

[+ See all authors and affiliations](#)

Science 16 Feb 2018;
Vol. 359, Issue 6377, pp. 725-726
DOI: 10.1126/science.359.6377.725

Article

Figures & Data

Info & Metrics

eLetters

 PDF

You are currently viewing the summary.

[View Full Text](#)



Summary

The booming field of artificial intelligence (AI) is grappling with a replication crisis, much like the ones that have afflicted psychology, medicine, and other fields over the past decade. Just because algorithms are based on code doesn't mean experiments are easily replicated. Far from it. **Unpublished codes and a sensitivity to training conditions have made it difficult for AI researchers to reproduce many key results.** That is leading to a new conscientiousness about research methods and publication protocols. Last week, at a meeting of the Association for the Advancement of Artificial Intelligence in New Orleans, Louisiana, reproducibility was on the agenda, with some teams diagnosing the problem—and one laying out tools to mitigate it.

» The replication and criticism movement is not about suppressing speculative research; rather, it's all about establishing science's fabled self-correcting nature
The world's most popular languages that the blog documentation hasn't been translated into »

Do differences between biology and statistics explain some of our diverging attitudes regarding criticism and replication of scientific claims?

Posted by Andrew

Last month we had much of a push that it's in ever-replicated, this possibly a frag (possibly associated) significant claim.

So we disagree on structures of ob-

The people point of view. Such people replicate a lot of work to do.

She obviously profile put they've done don't have you can do.

I agree, and I think journals, or my statistics: Biggest the bottom I think counts for more.

motivation for people of lower ranks to make a reputation by tangling with someone higher up. Put it all together and you have some toxic politics, much different than what you'll see in a flatter field such as statistics.

Or so I speculate.

<http://andrewgelman.com/>



SEARCH

RECENT COMMENTS

- » Samerria Daniels on Why is the replication crisis centered on social psychology?
- » Keith O'Rourke on What killed alchemy?
- » zbicylest on Why is the replication crisis centered on social psychology?
- » Anonymous on What killed alchemy?
- » Jeff Valentine on Why is the replication crisis centered on social psychology?
- » Jeff Valentine on Why is the replication crisis centered on social psychology?
- » Anonymous on What killed alchemy?
- » Samerria Daniels on Why is the replication crisis centered on social psychology?
- » Anonymous on Why is the replication crisis centered on social psychology?
- » Keith O'Rourke on Why is the replication crisis centered on social psychology?
- » Keith O'Rourke on Why is the replication crisis centered on social psychology?
- » jkirkdeau on Why is the replication crisis centered on social psychology?

If I say *just trust me* and I'm wrong,
I'm **untrustworthy**.

If I say *here's my work* and it's wrong,
I'm **honest**, human,
and serving scientific progress.

When you try to reproduce a result,
but an essential method is missing



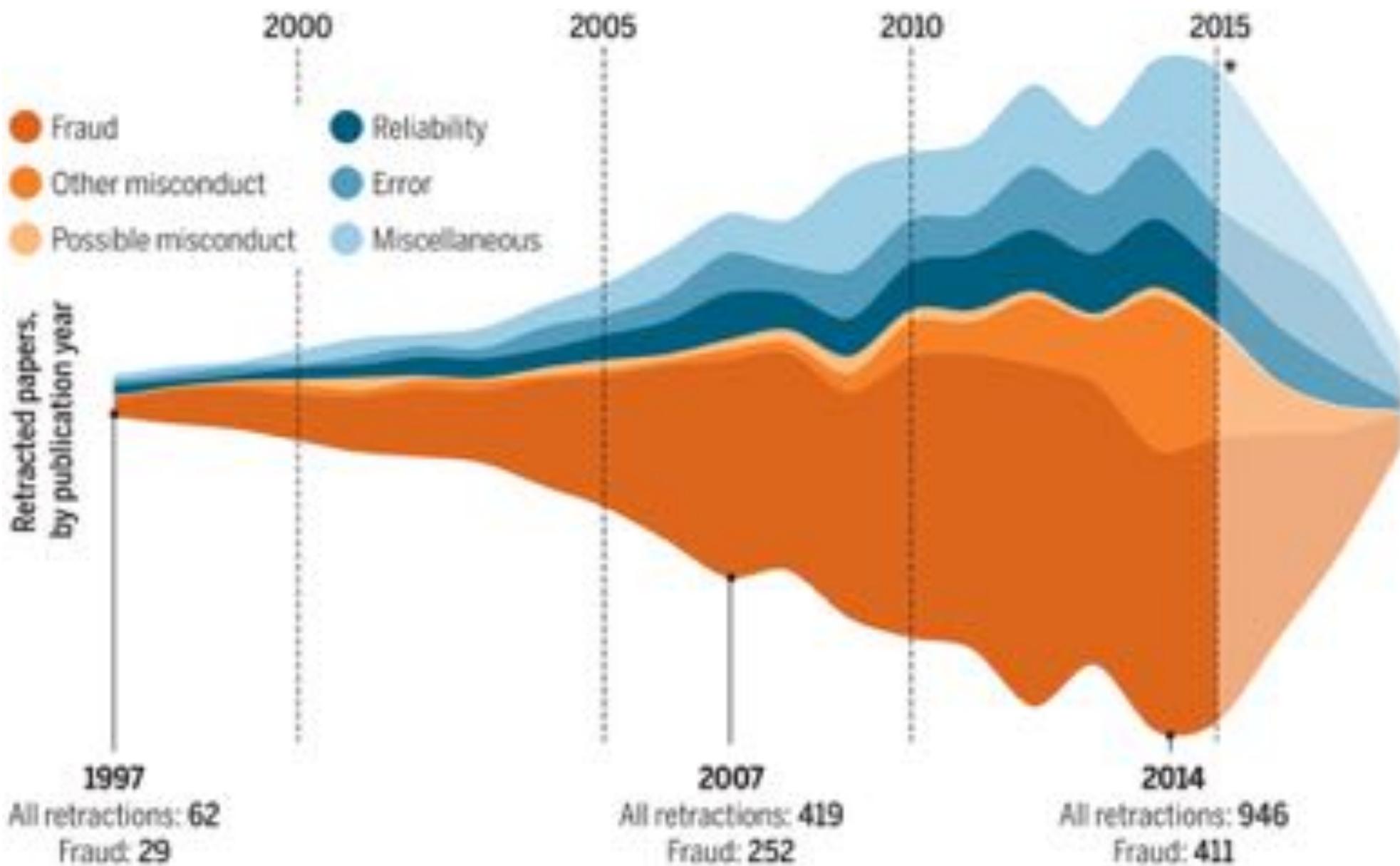
Found on twitter

DERIVING CHEMOSENSITIVITY FROM CELL LINES: FORENSIC BIOINFORMATICS AND REPRODUCIBLE RESEARCH IN HIGH-THROUGHPUT BIOLOGY

BY KEITH A. BAGGERLY¹ AND KEVIN R. COOMBES²

University of Texas

High-throughput biological assays such as microarrays let us ask very detailed questions about how diseases operate, and promise to let us personalize therapy. Data processing, however, is often not described well enough to allow for exact reproduction of the results,





No witchhunt anywhere – Tim Errington's talk on the Reproducibility Project: Cancer Biology

★★★★★ 3 Votes



"How reproducible is cancer biology?" stood in bold letters on a poster I had designed to advertise a talk by Tim Errington, one of the leaders of the [Reproducibility Project: Cancer Biology](#) (RP:CB), in Cambridge a few weeks ago. And I had told everyone "Come and learn who's good, who's bad and who's ugly in cancer research!"

Cus

Journal of Visual Experiments

The screenshot shows the homepage of the Journal of Visual Experiments (JoVE) website. At the top, the JoVE logo is on the left, followed by a search bar with "Advanced" and a magnifying glass icon. To the right are "START A TRIAL" and "LOG IN" buttons. Below the header, there are five navigation tabs: "ABOUT JoVE", "FOR LIBRARIANS", "PUBLISH" (which is highlighted in dark blue), "VIDEO JOURNAL", and "SCIENCE EDUCATION". The main content area features a large video thumbnail on the left showing a close-up of a brain slice being tested with electrodes. To the right of the video, the tagline "WE CHANGE THE WAY SCIENCE IS DONE" is displayed in white text on a dark blue background. On the far left of the main content area, there is a "NOW PLAYING" section with a vertical list of five video thumbnails and their titles. The first video is currently selected.

NOW PLAYING

- Making, Testing, and Using Potassium Ion Selective Microelectrodes in Tissue Slices of Adult Brain
- Using a Fluorescence Microscope to Visualize Axon Pathfinding
- Measuring the Mechanical Properties of Neuronal Axons Using a Cross-Linear Viscoelasticity Method
- Visualizing Axon Pathfinding in Three-Dimensional Space Using Confocal Microscopy
- Establishing the Axonal Pathway of Dendrites in the Cerebral Cortex by Injecting Tracer Dyes

VIDEO JOURNAL

WE CHANGE
THE WAY SCIENCE
IS DONE