

Ginkgo — Interactive analysis and quality assessment of single-cell CNV data

 @Ty_Garvin

Tyler Garvin, Robert Aboukhalil, Jude Kendall, Timour Baslan,
Gurinder S. Atwal, Jim Hicks, Michael Wigler, Michael C. Schatz



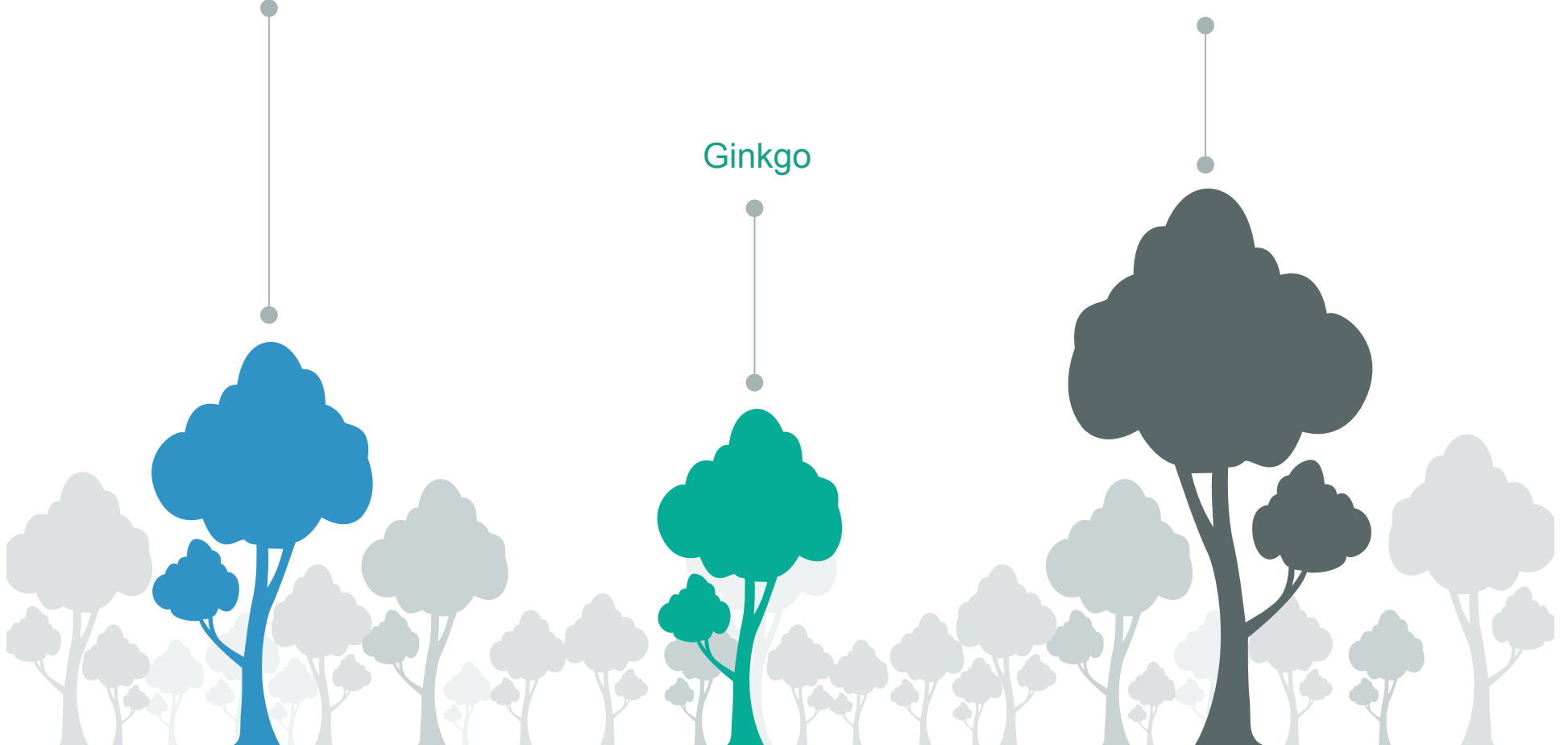
Cold Spring Harbor Laboratory

Outline

Introduction

Ginkgo

Comparison of
WGA methods



Outline

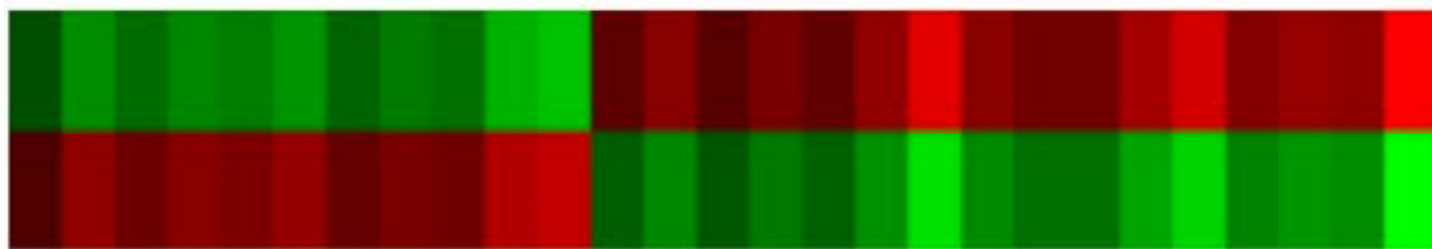
Introduction

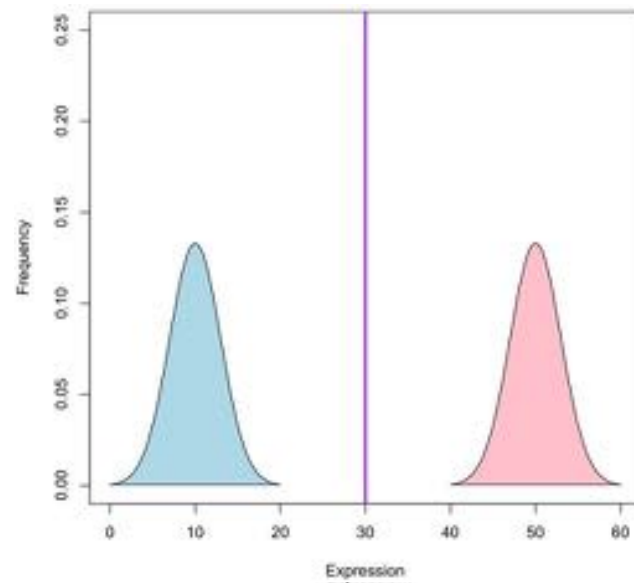
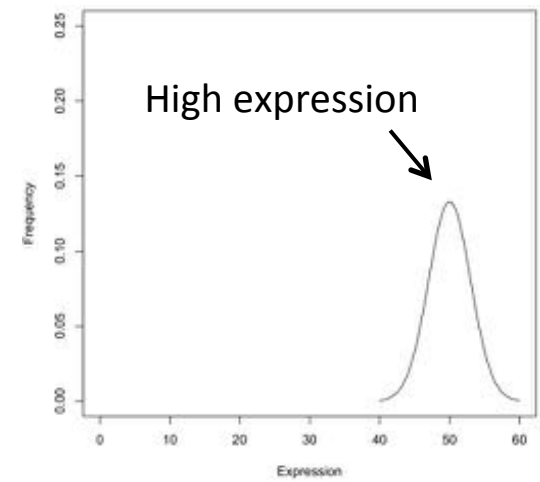
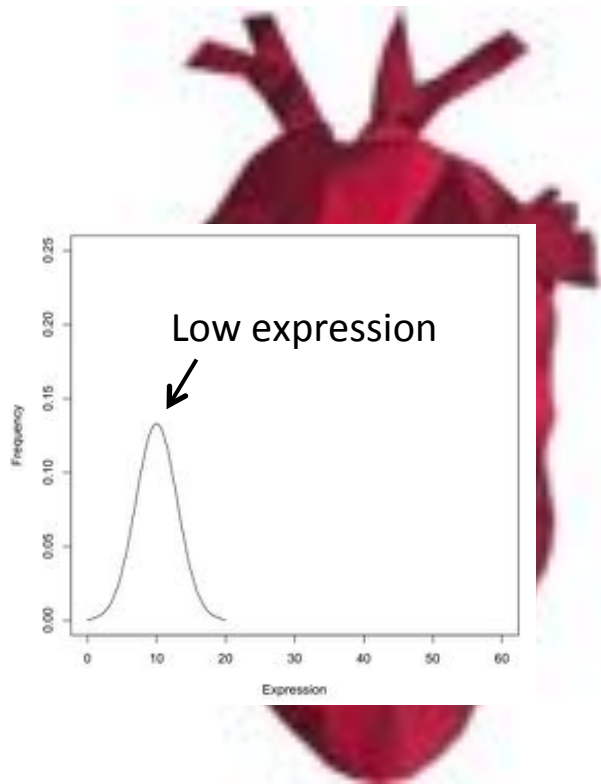
Comparison of
WGA methods

Ginkgo

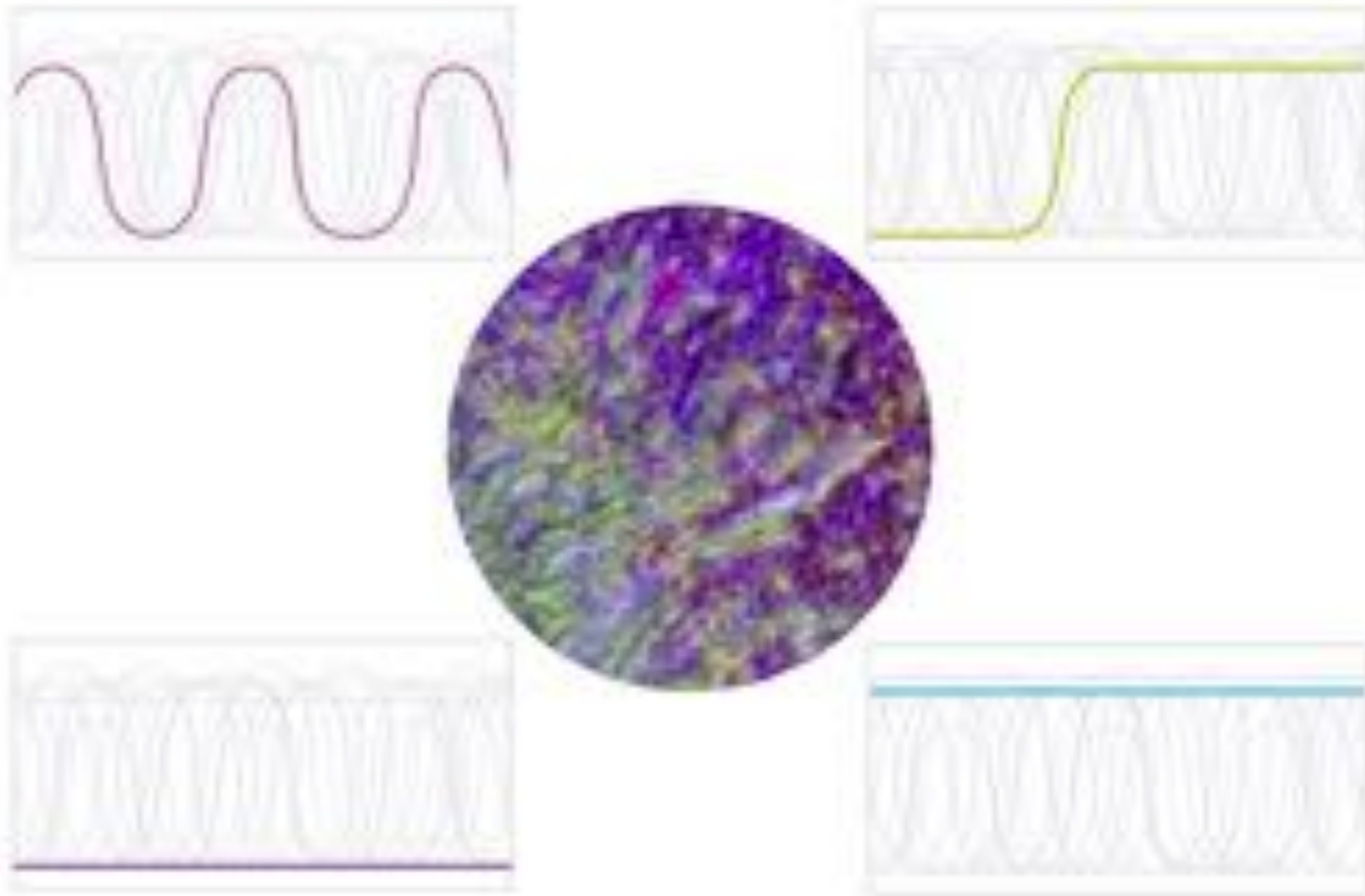


Why should we use single-cell sequencing over bulk sequencing?





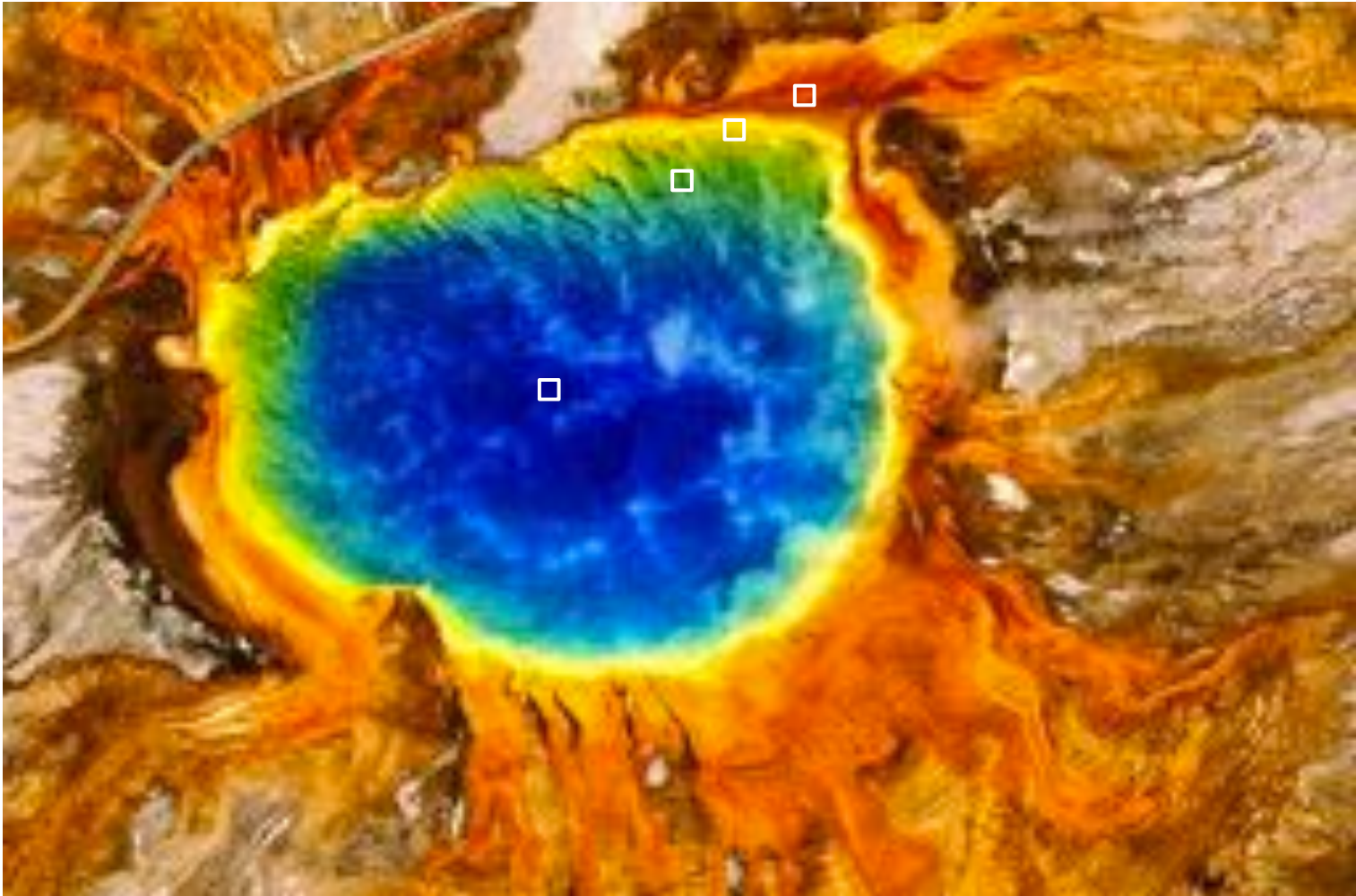
Single cell expression profiles



Single cell research. Illumina.

Heterogeneity

Metagenomics



Single cell research. Illumina.

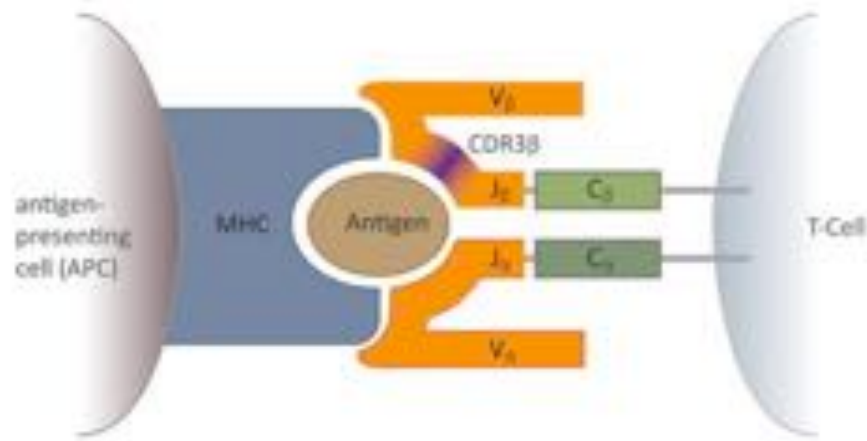
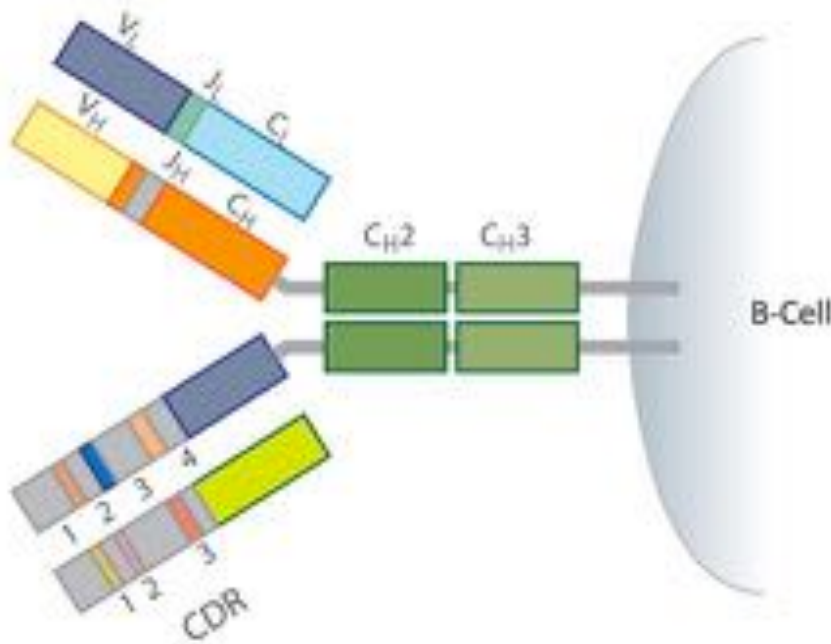


Microbiome

- Human cells
 - Bacteria
 - Fungi
 - Virus
-
- Temporal dynamics
 - Diet-dependent dynamics
 - Stress-related dynamics

Immunology

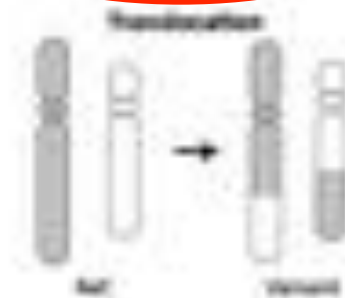
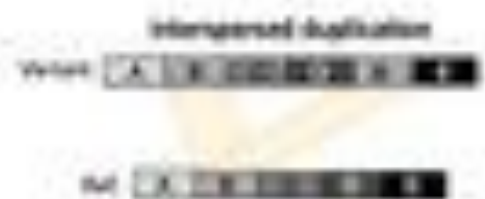
- Massive diversity rivaled only by germ cells
- Somatic recombination



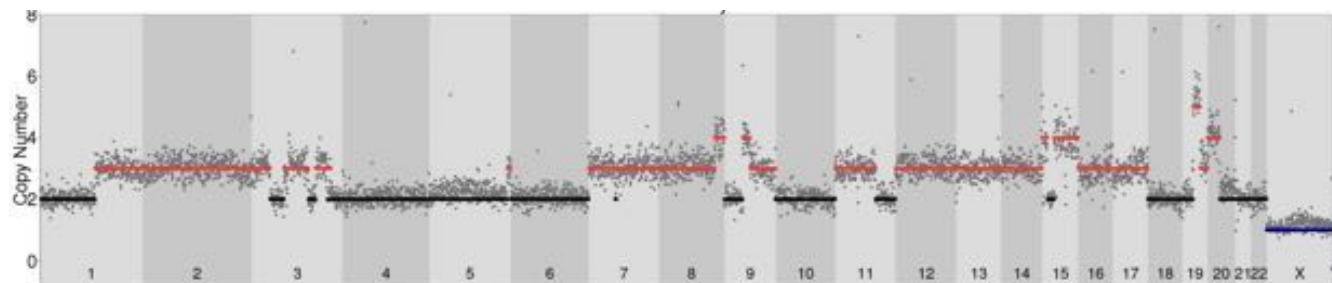
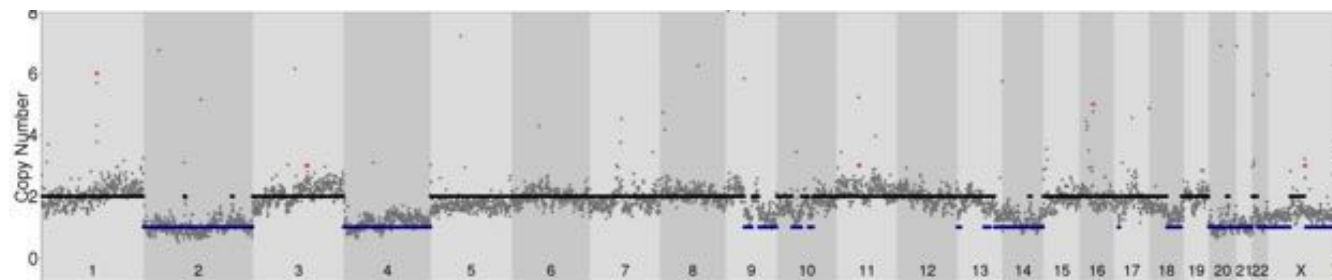
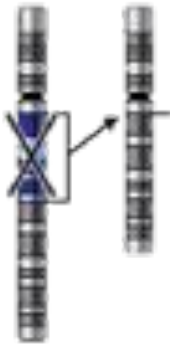
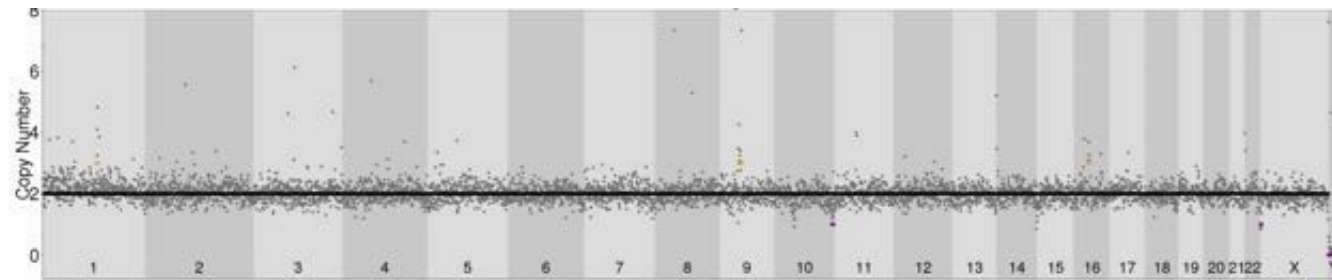
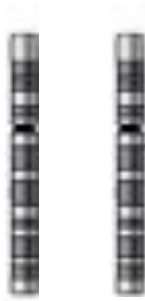
- B cells – antibody generation
- T cells – antigen response

What is structural variation?

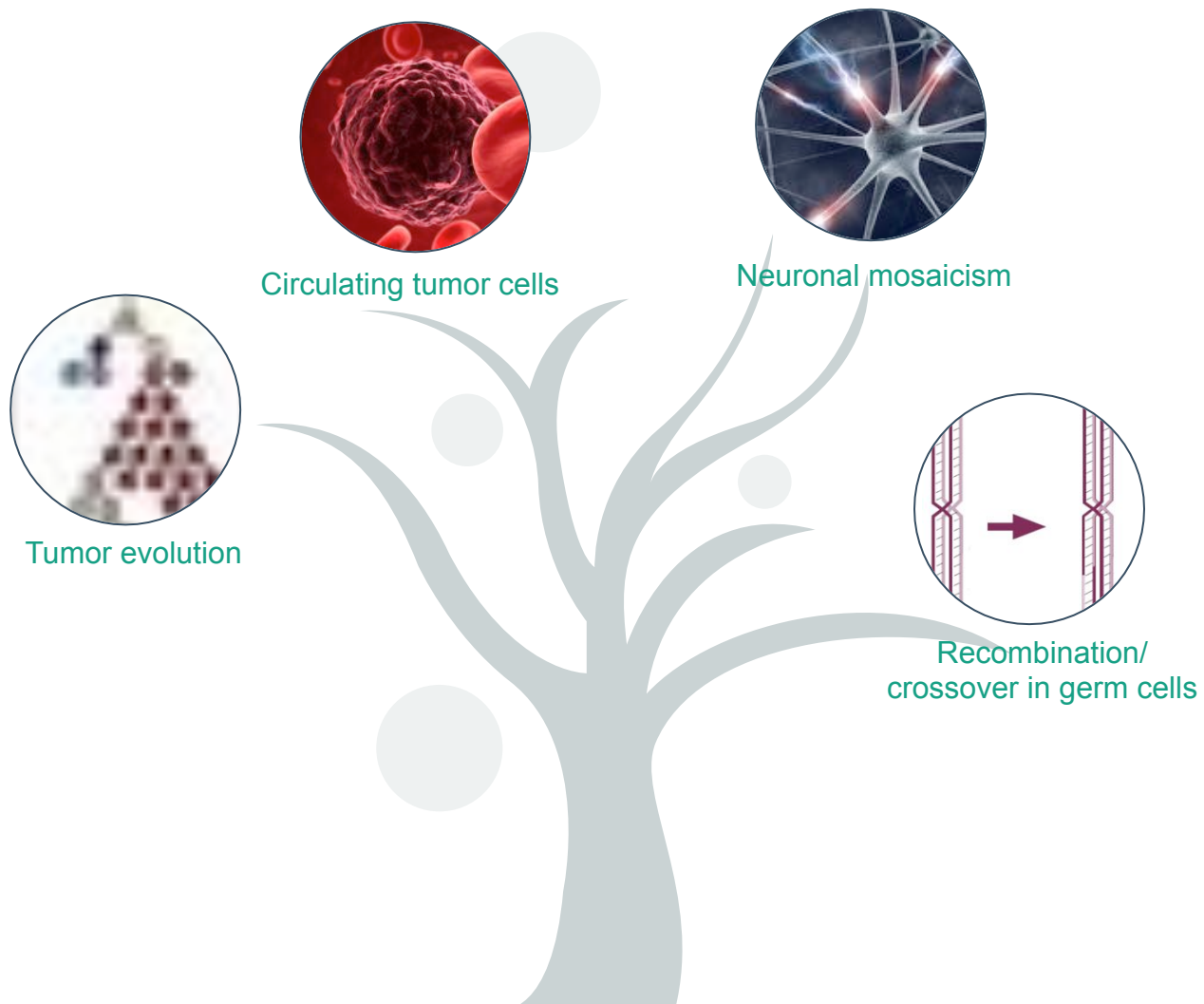
Difference in copy number, orientation, or location of any genomic sequence over 50 bp in size



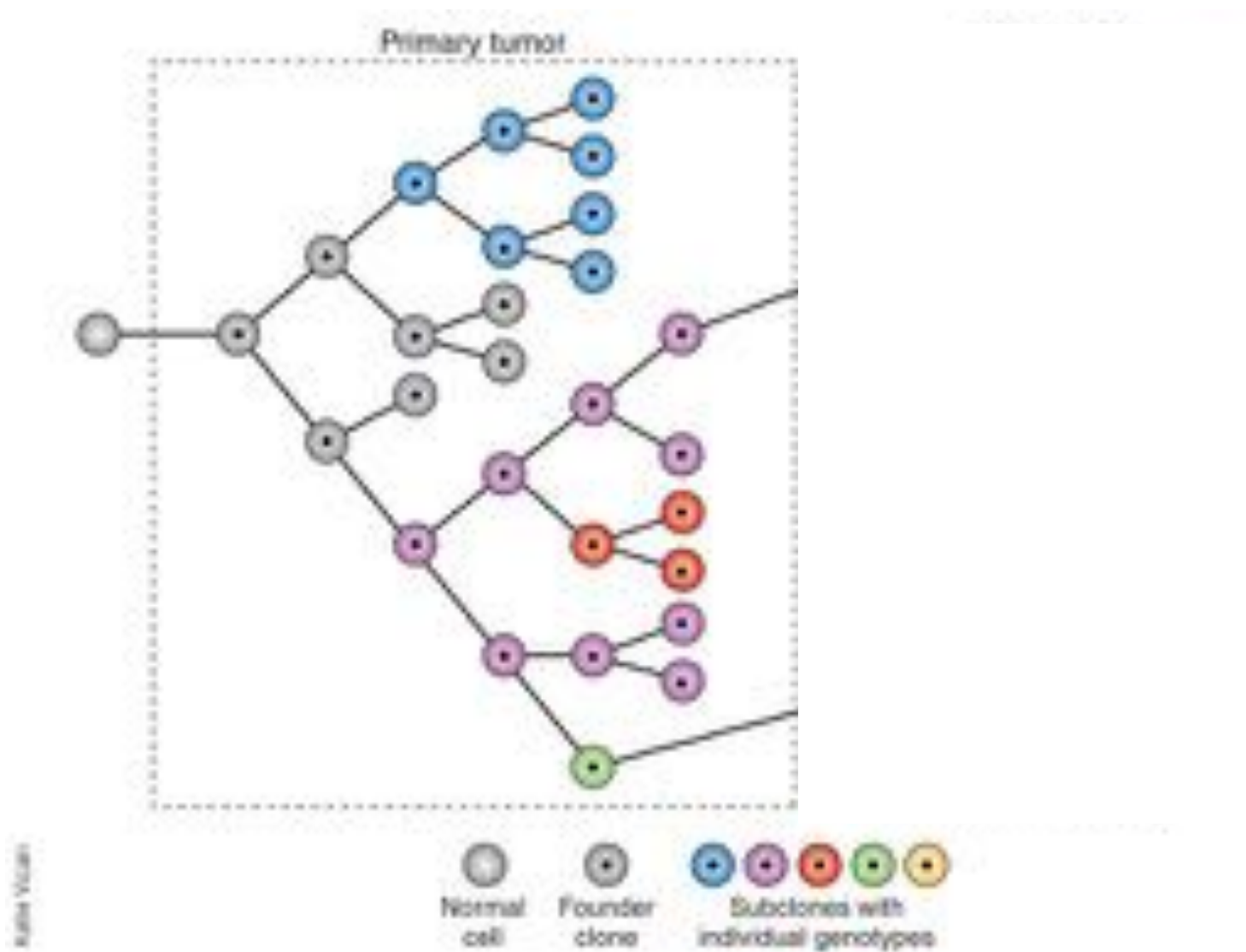
What are CNVs?



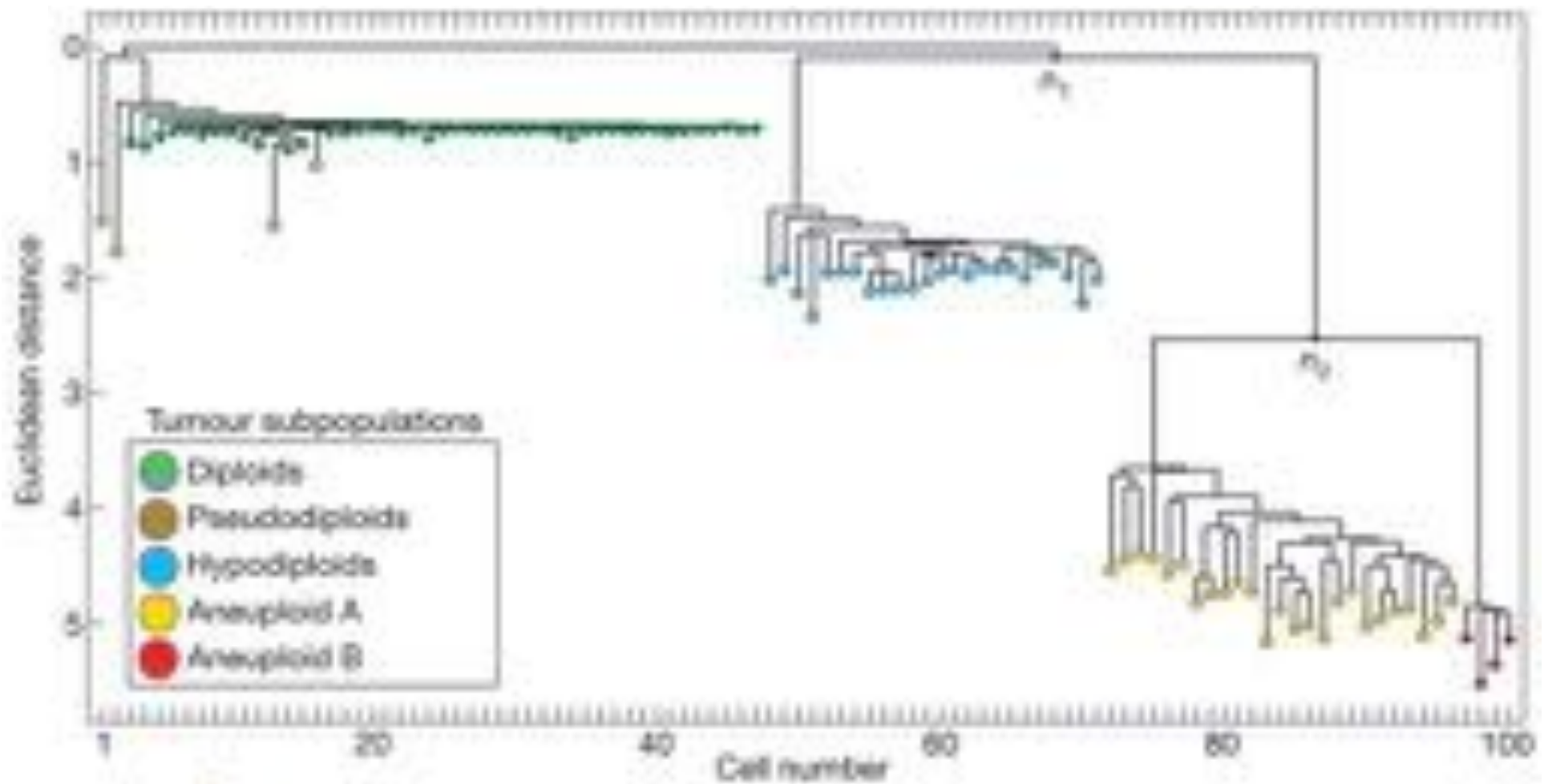
Single-cell sequencing for CNV analysis



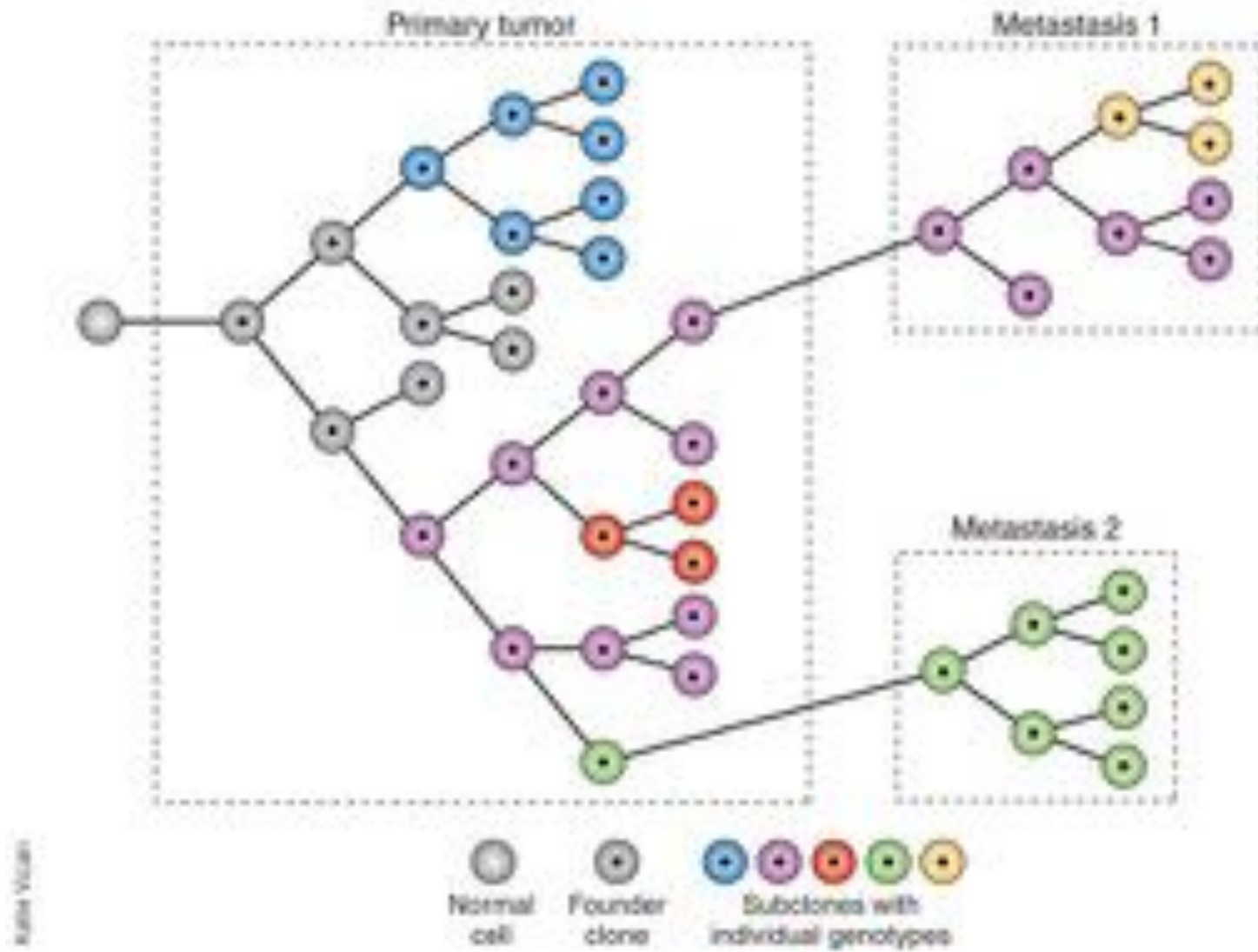
Clonal evolution in tumors



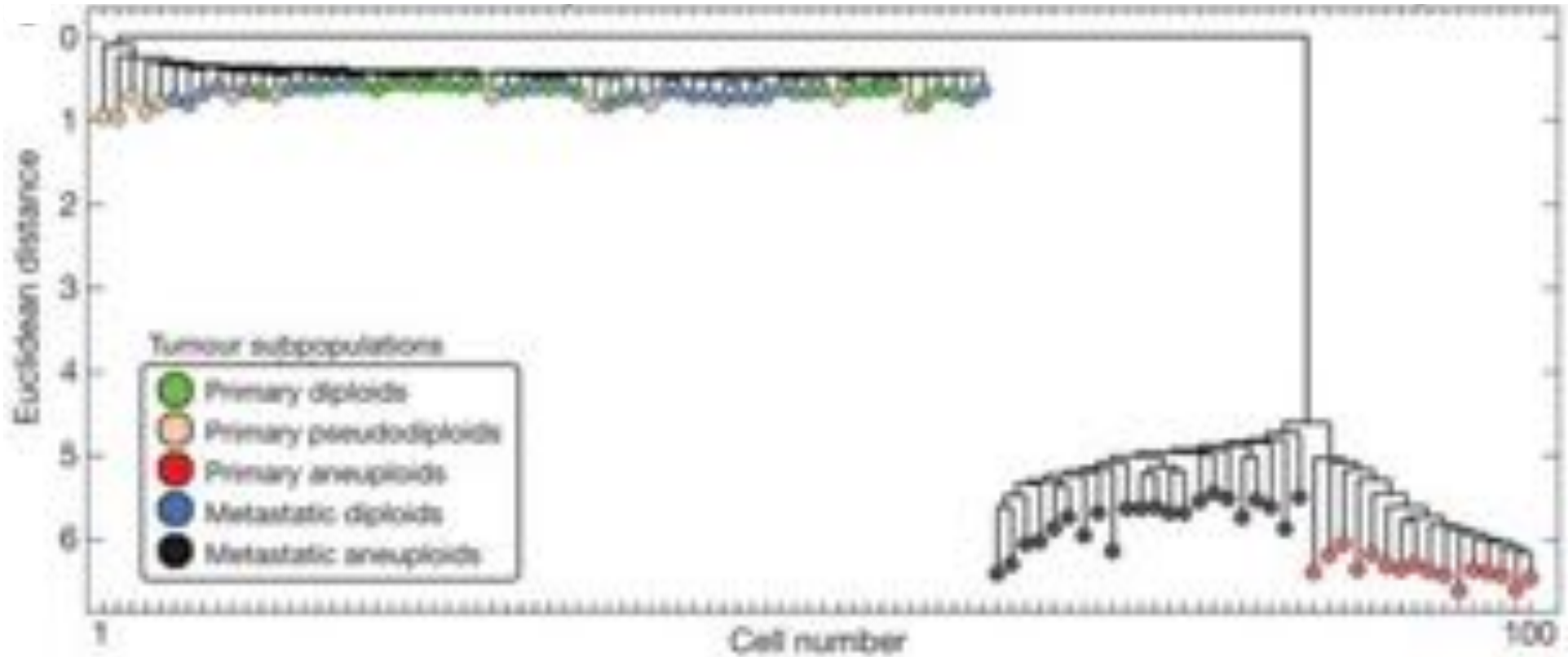
Clonal evolution in tumors



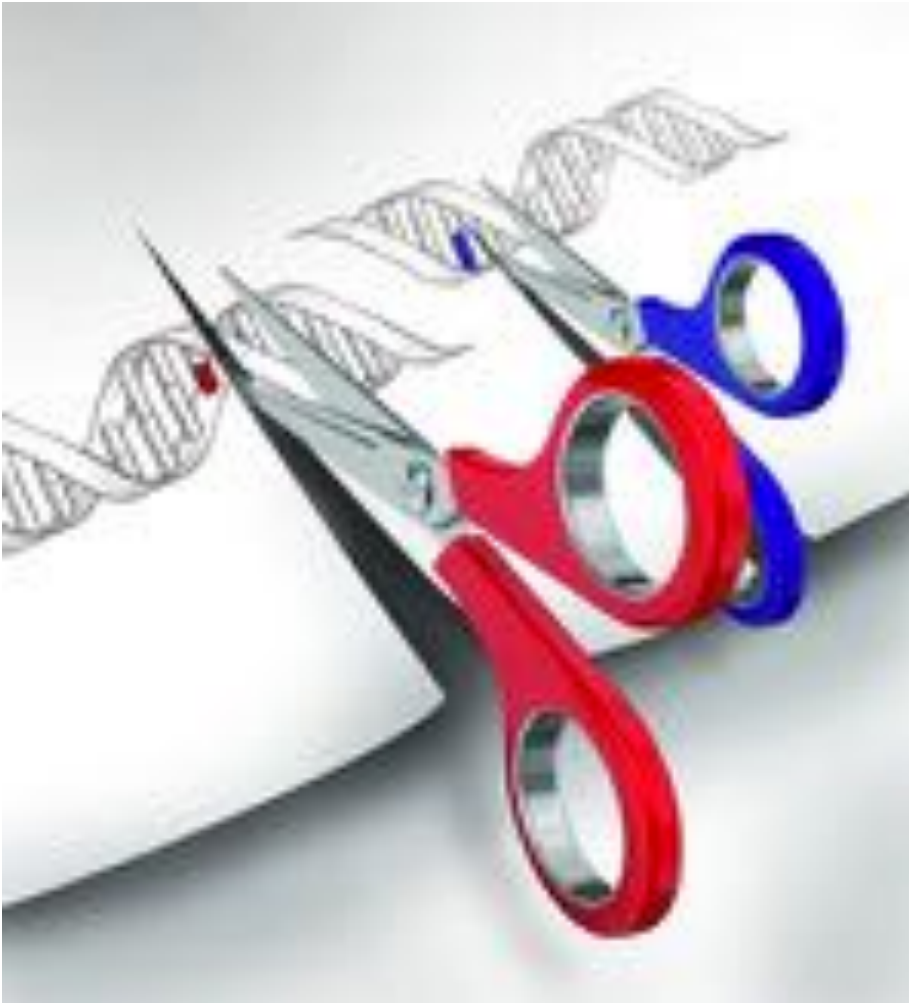
Cancer metastasis



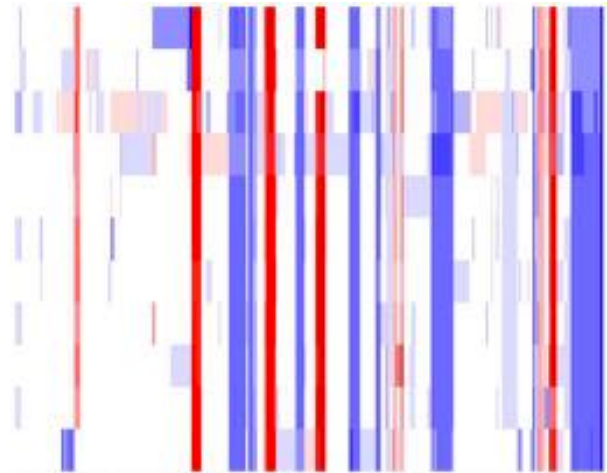
Cancer metastasis



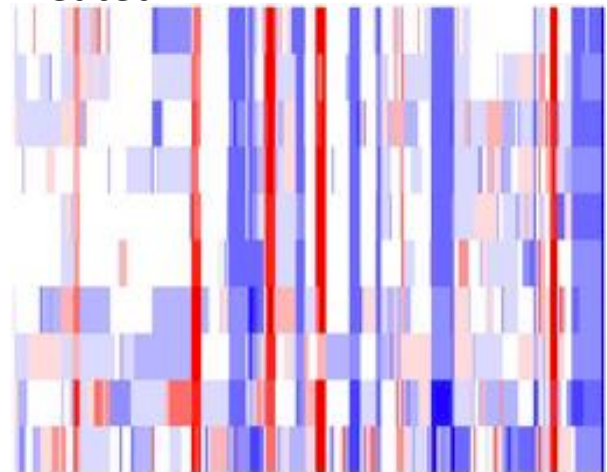
DNA repair in cancer



Untreated

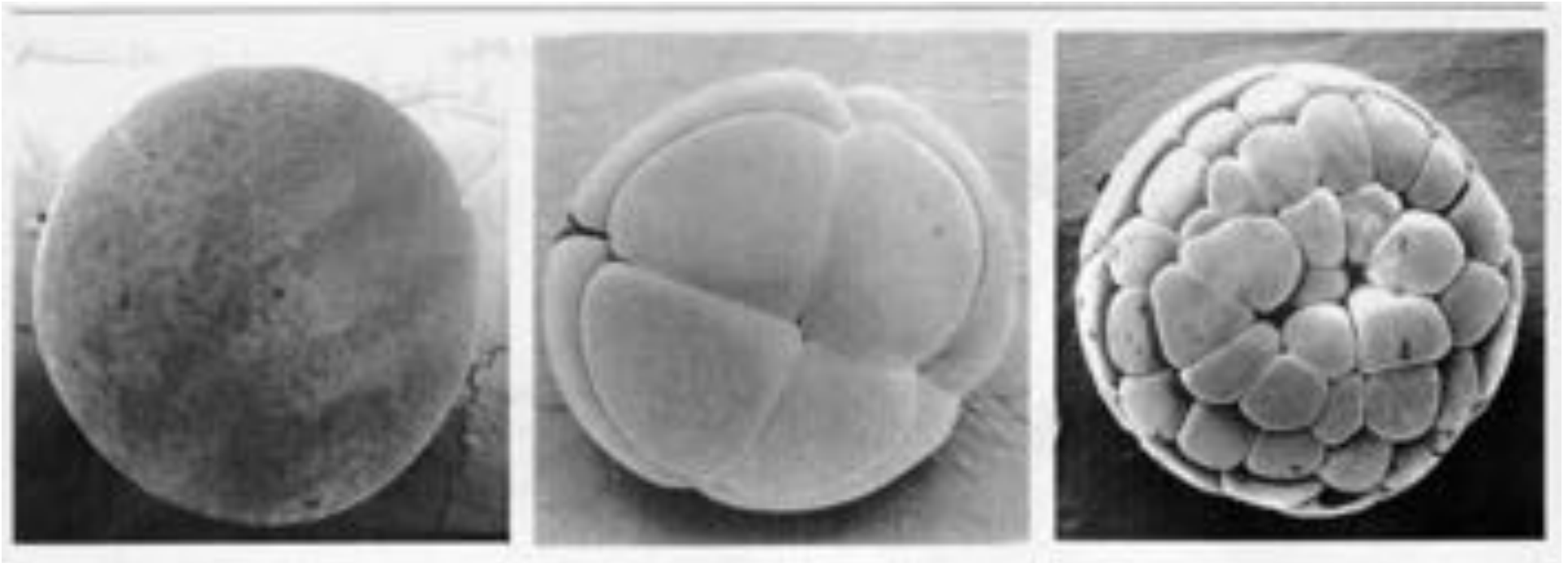


Treated

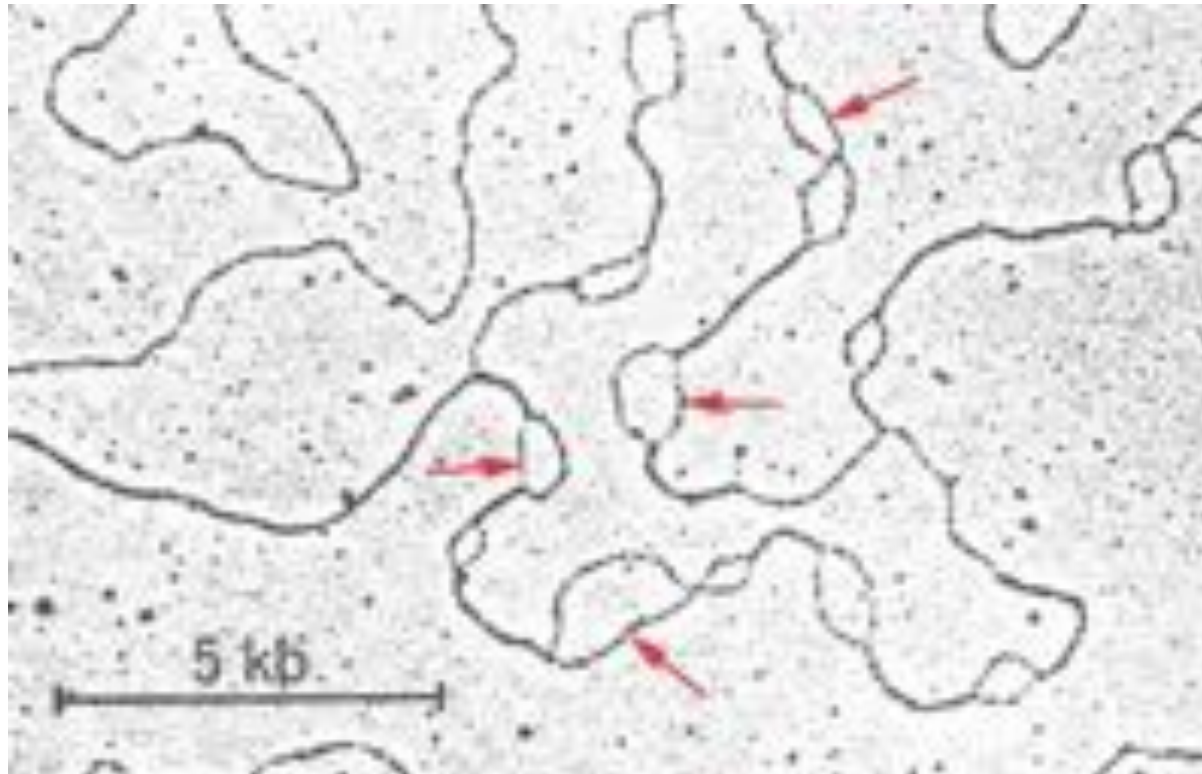


← Genome →

In-vitro fertilization

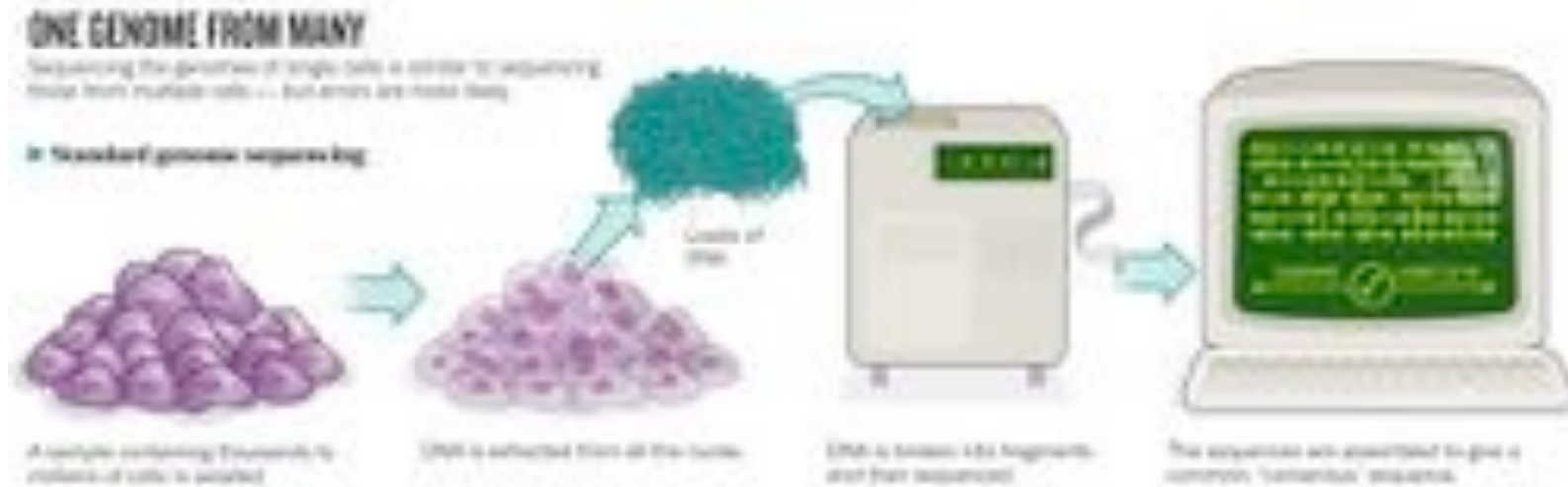


DNA replication & cell cycle

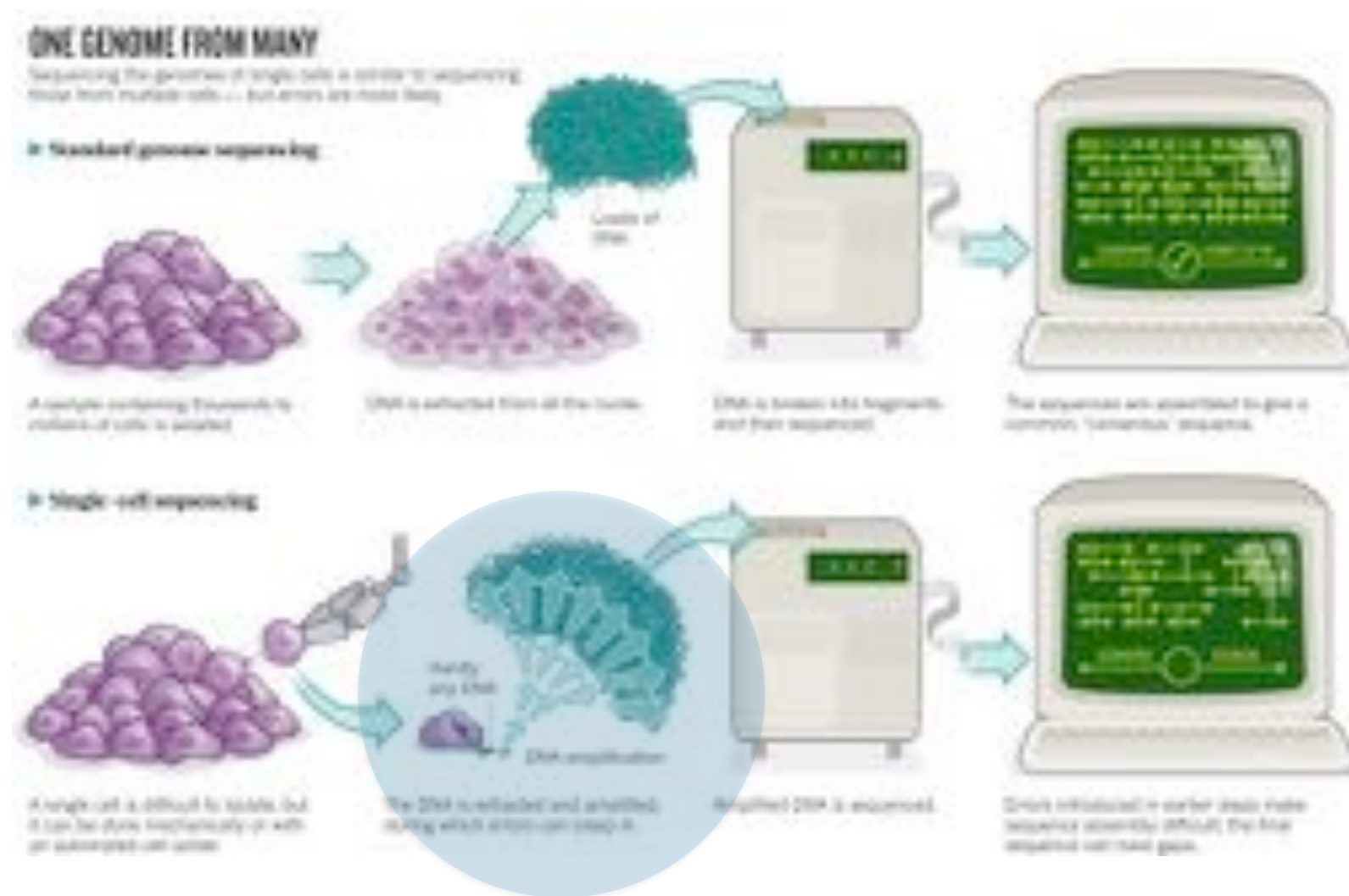


late vs. early replicating regions

Single-cell vs. bulk sequencing



Single-cell vs. bulk sequencing

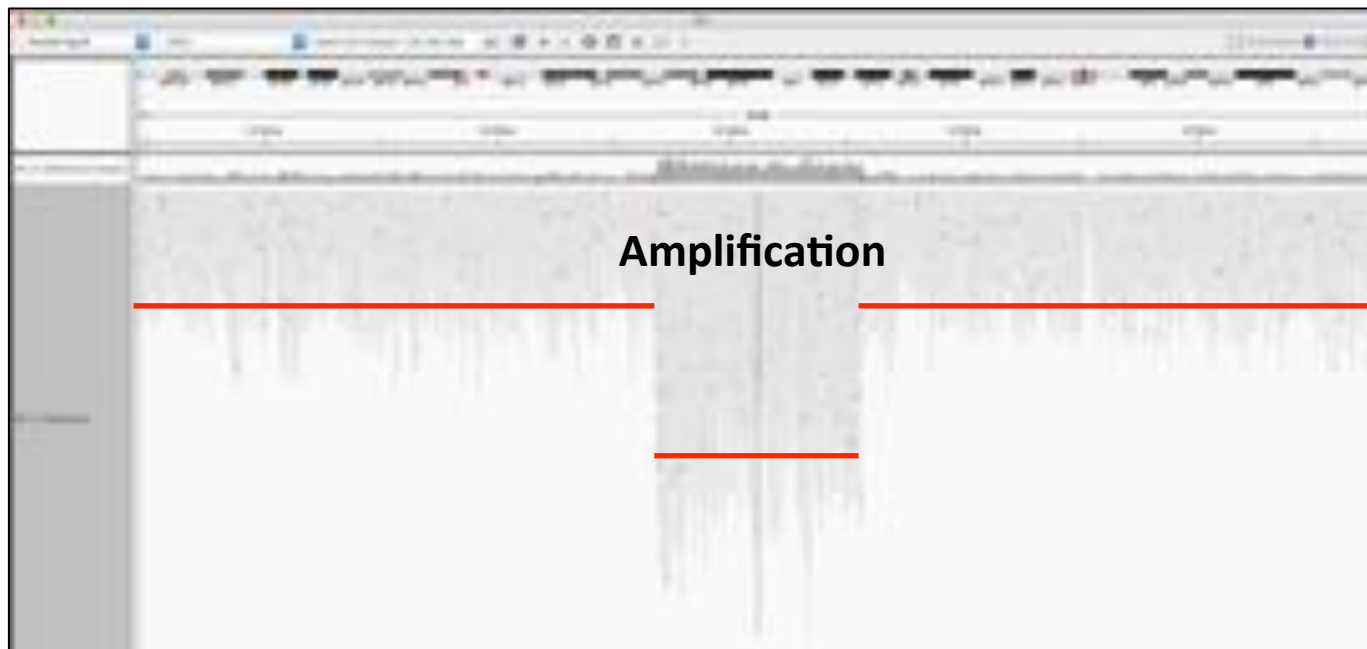
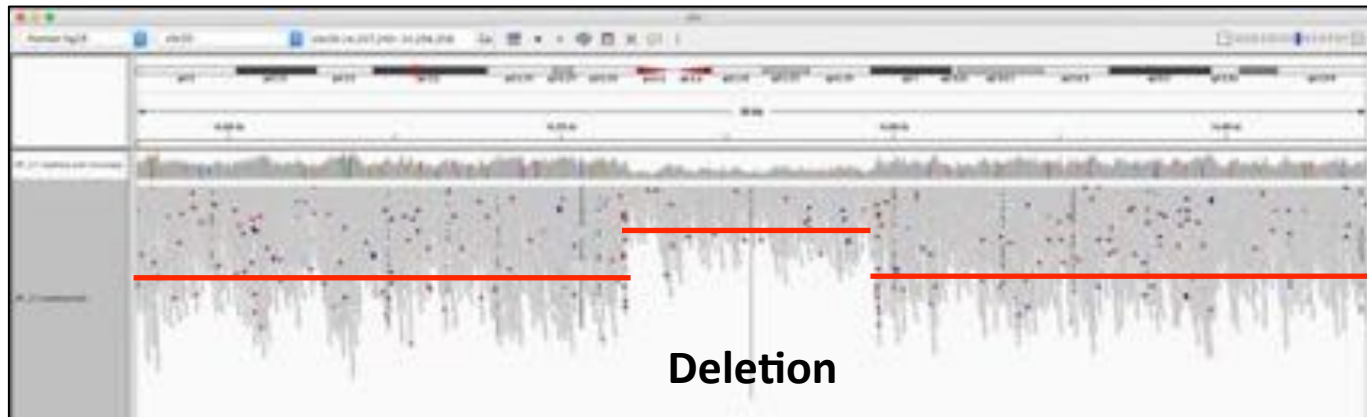


Whole Genome Amplification



- 1) MDA: Multiple displacement amplification
- 2) DOP-PCR: Degenerate oligonucleotide-primed PCR
- 3) MALBAC: Multiple annealing and looping-based amplification

Identifying CNVs: bulk vs. single-cell



Identifying CNVs: bulk vs. single-cell



Low coverage allows us to study copy-number variants

<1X coverage, often <0.1X

Copy-number variant analysis



Low coverage allows us to study copy-number variants

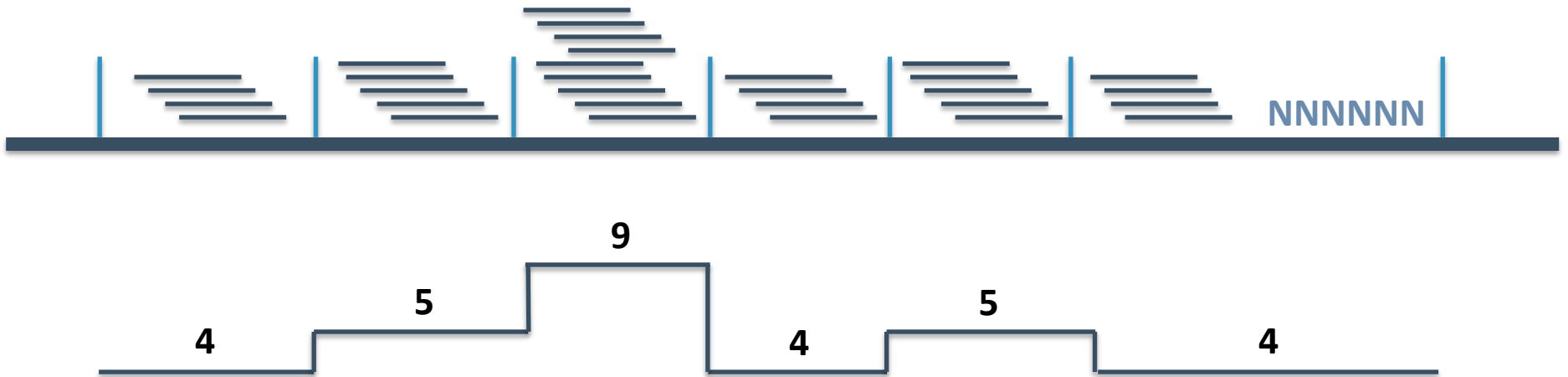
<1X coverage, often <0.1X

Copy-number variant analysis



Divide genome into “bins” with ~50 – 100 reads / bin

Copy-number variant analysis

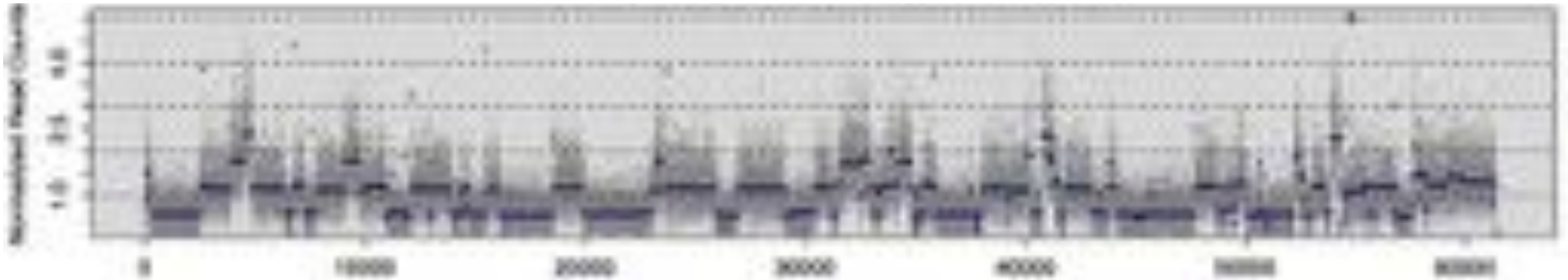


Copy-number variant analysis

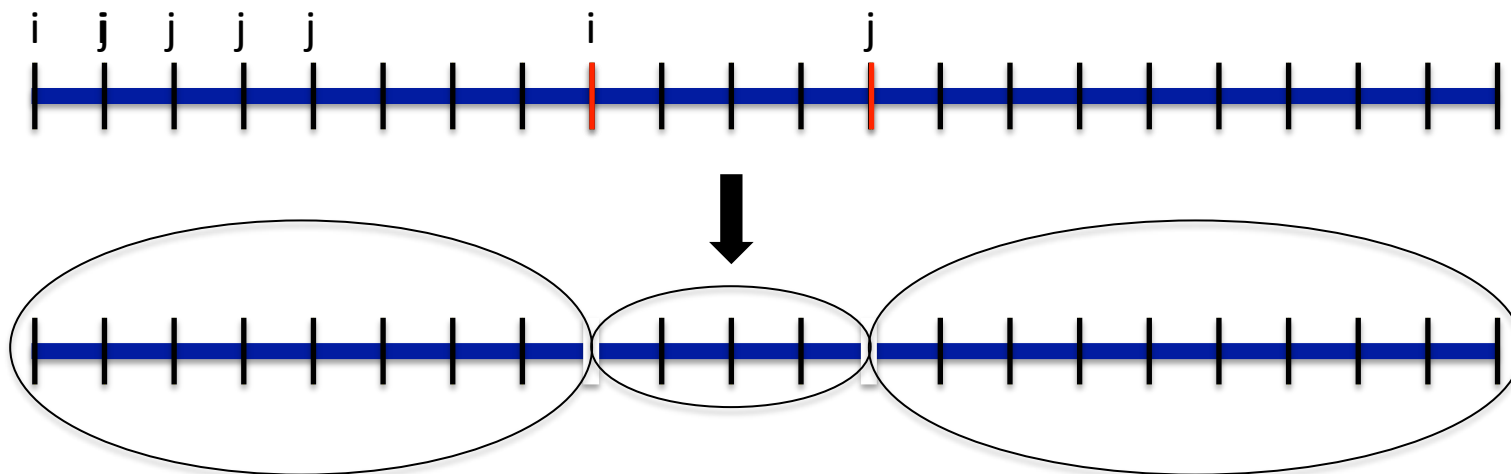


Circular Binary Segmentation (CBS)
to reduce noise in data

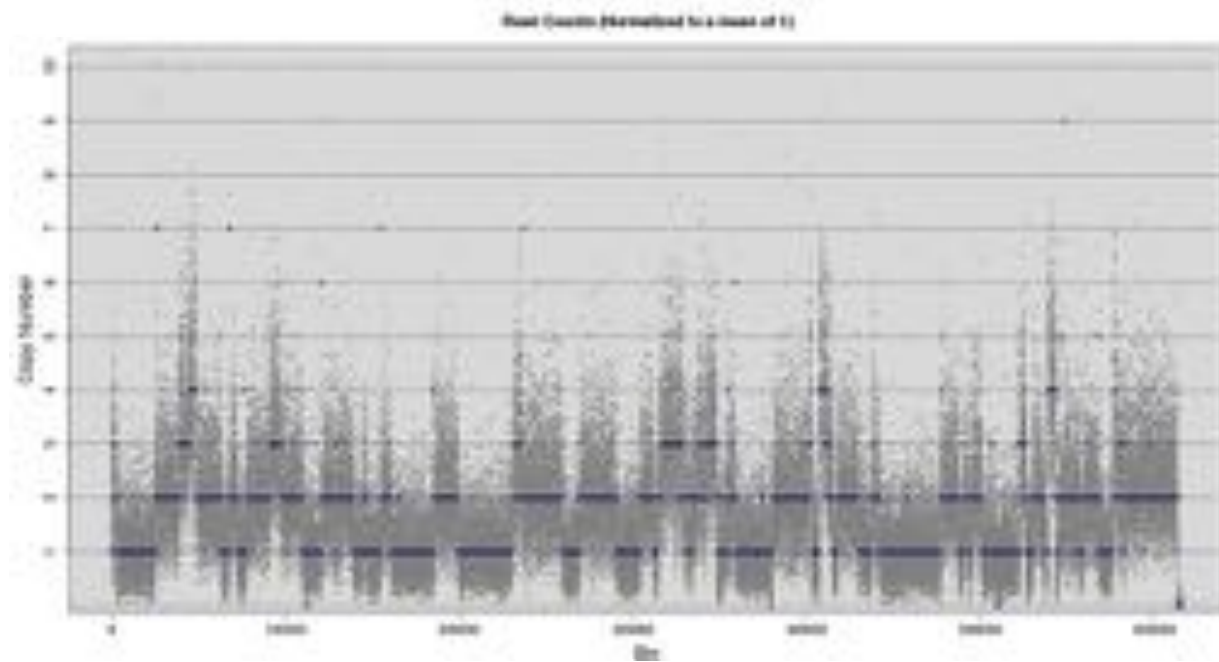
Segmentation



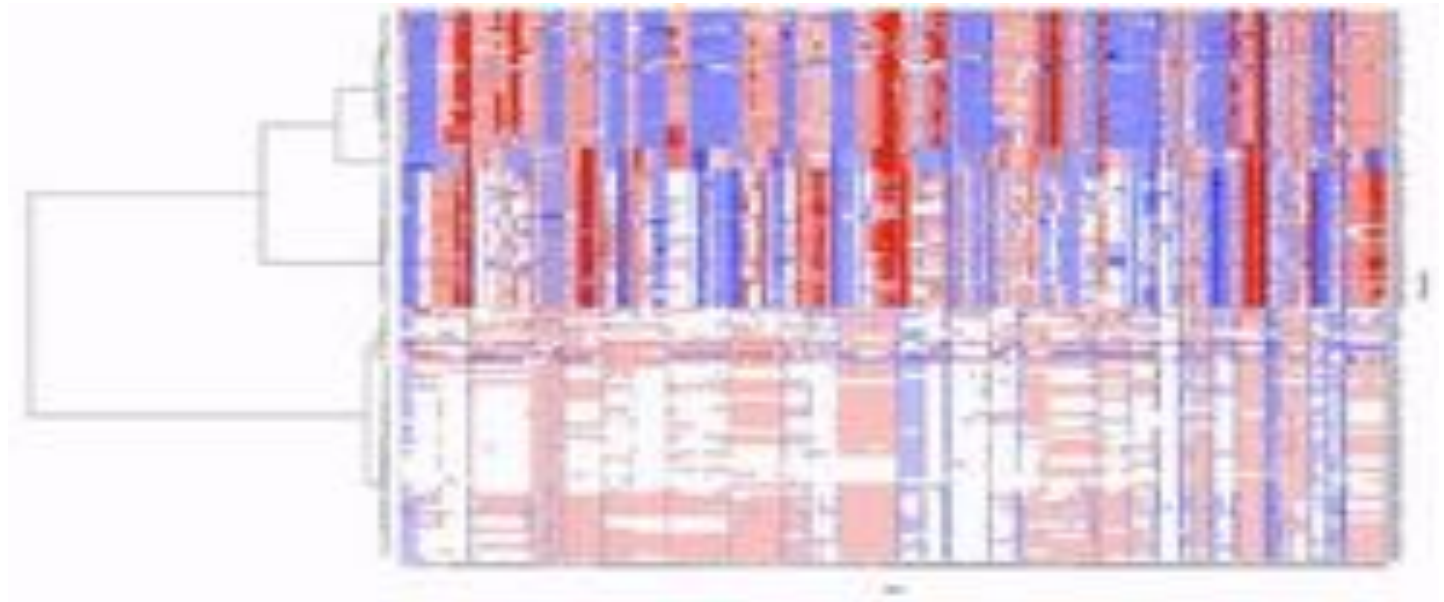
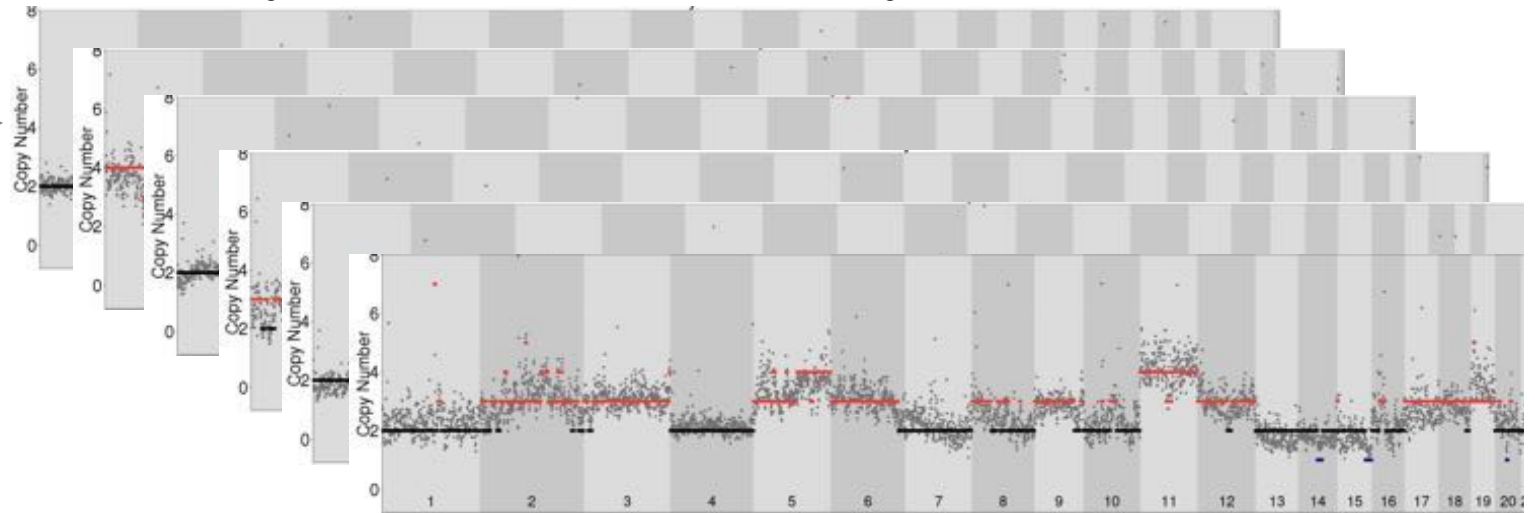
Circular Binary Segmentation (CBS)



Copy-number variant analysis



Copy-number variant analysis

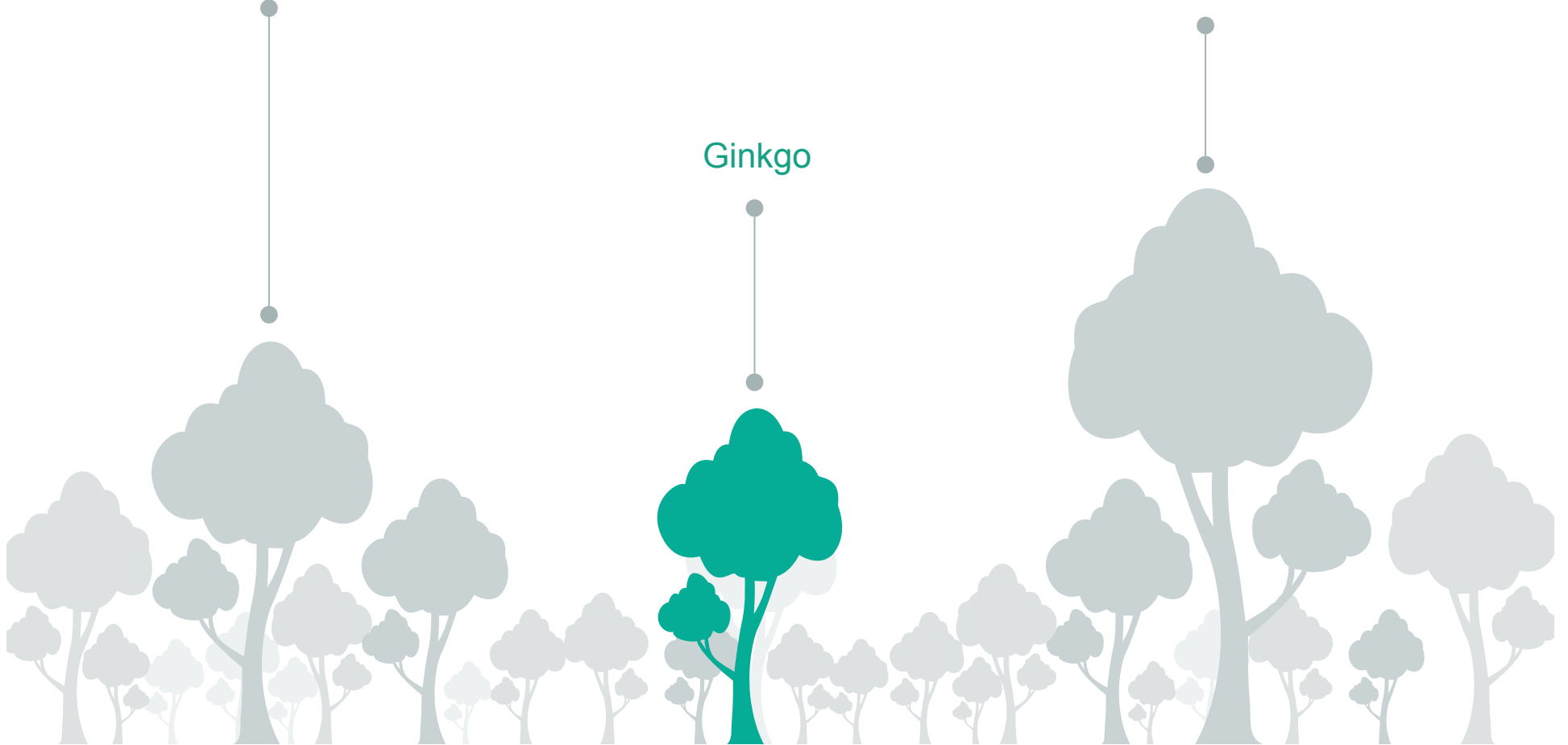


Outline

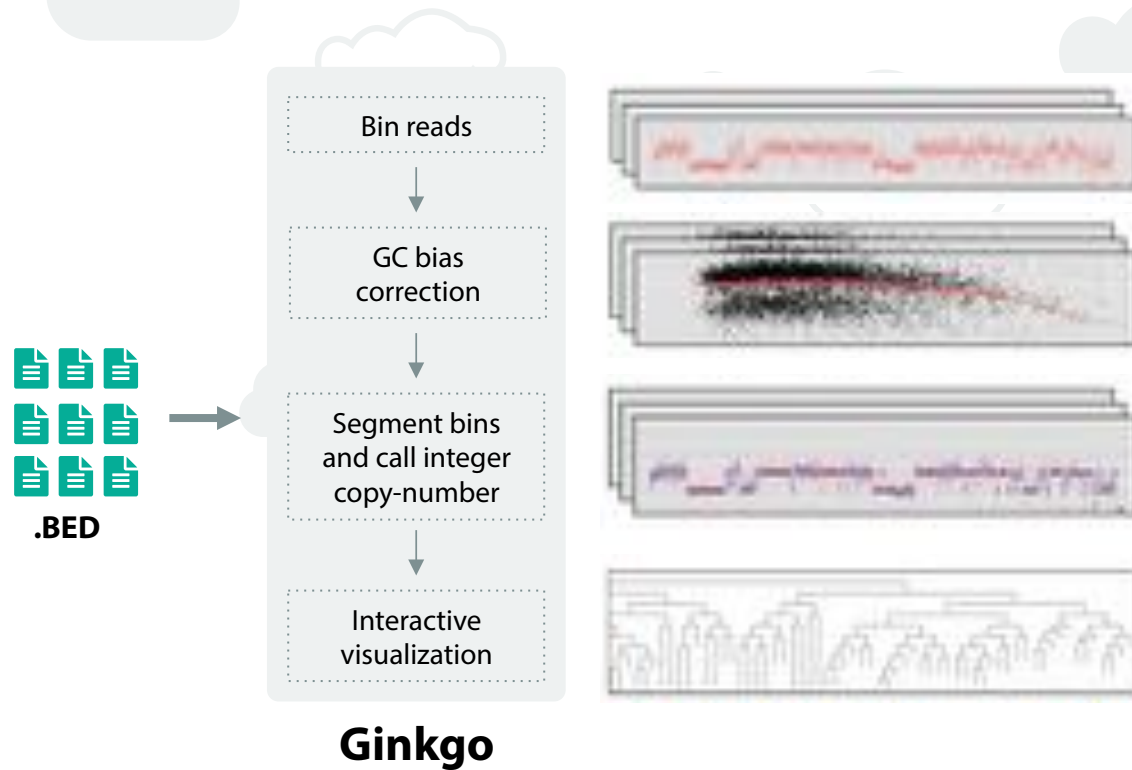
Introduction

Comparison of
WGA methods

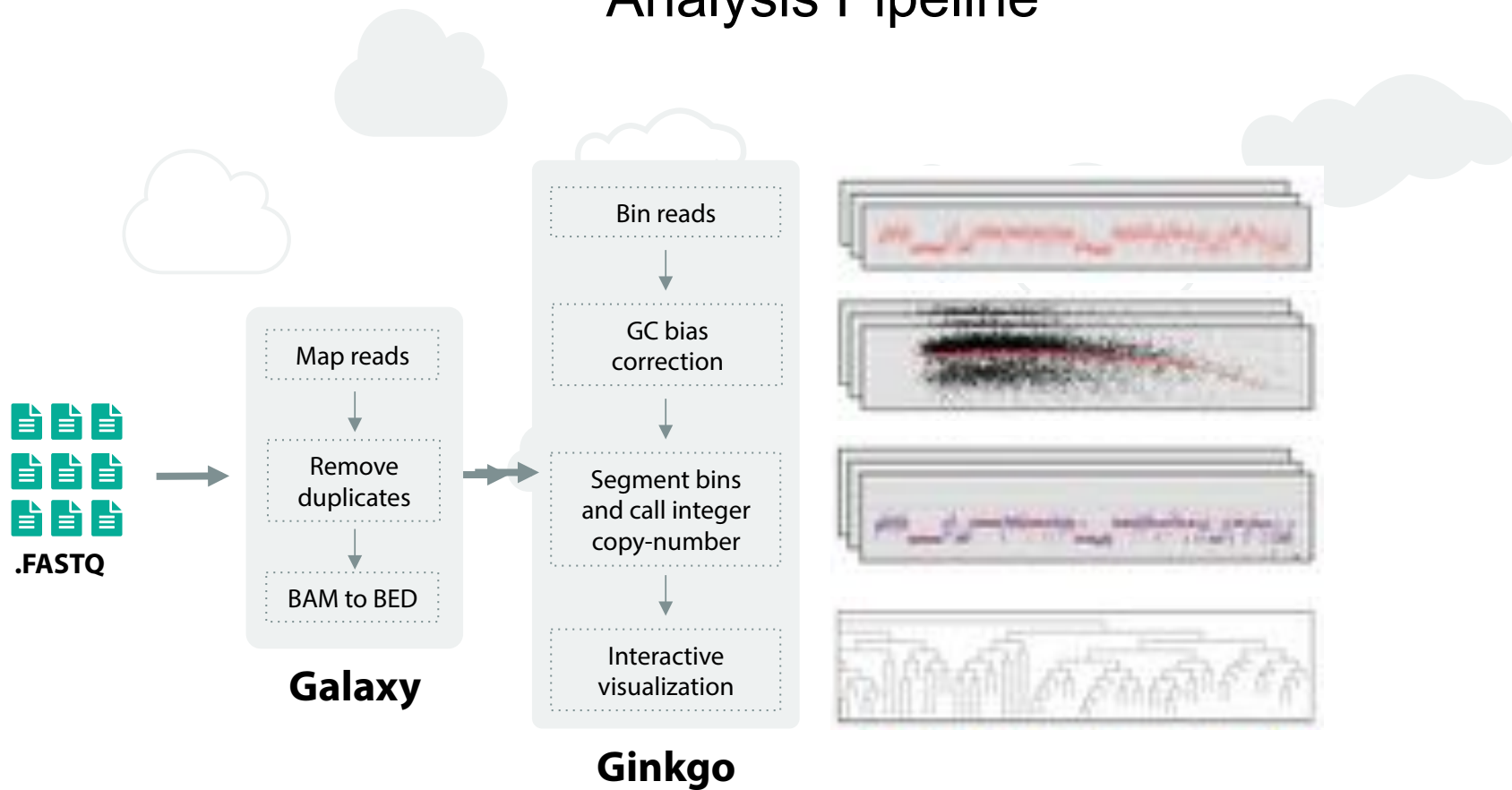
Ginkgo



Analysis Pipeline



Analysis Pipeline

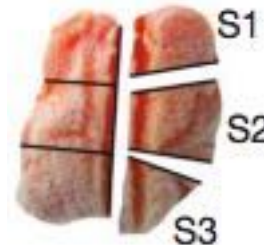
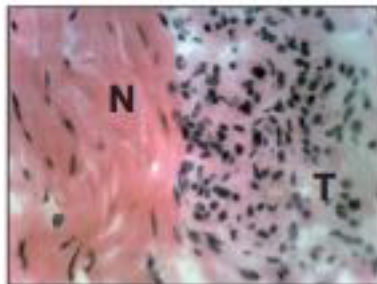


Ginkgo Demo

Tumour evolution inferred by single-cell sequencing

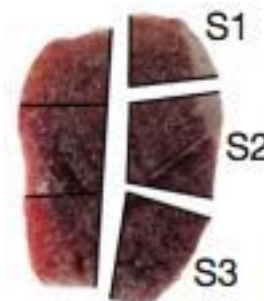
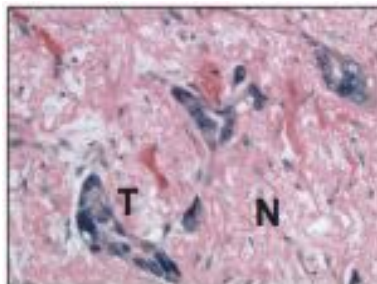
Nicholas Navin^{1,2}, Jude Kendall¹, Jennifer Tiroe¹, Peter Andrews¹, Linda Rodgers¹, Jeanne McIndoo¹, Kerry Cook¹, Asya Stepansky¹, Dan Levy¹, Diane Esposito¹, Lakshmi Muthuswamy³, Alex Krasnitz², W. Richard McCombie¹, James Hicks¹ & Michael Wigler¹

Breast
Tumor

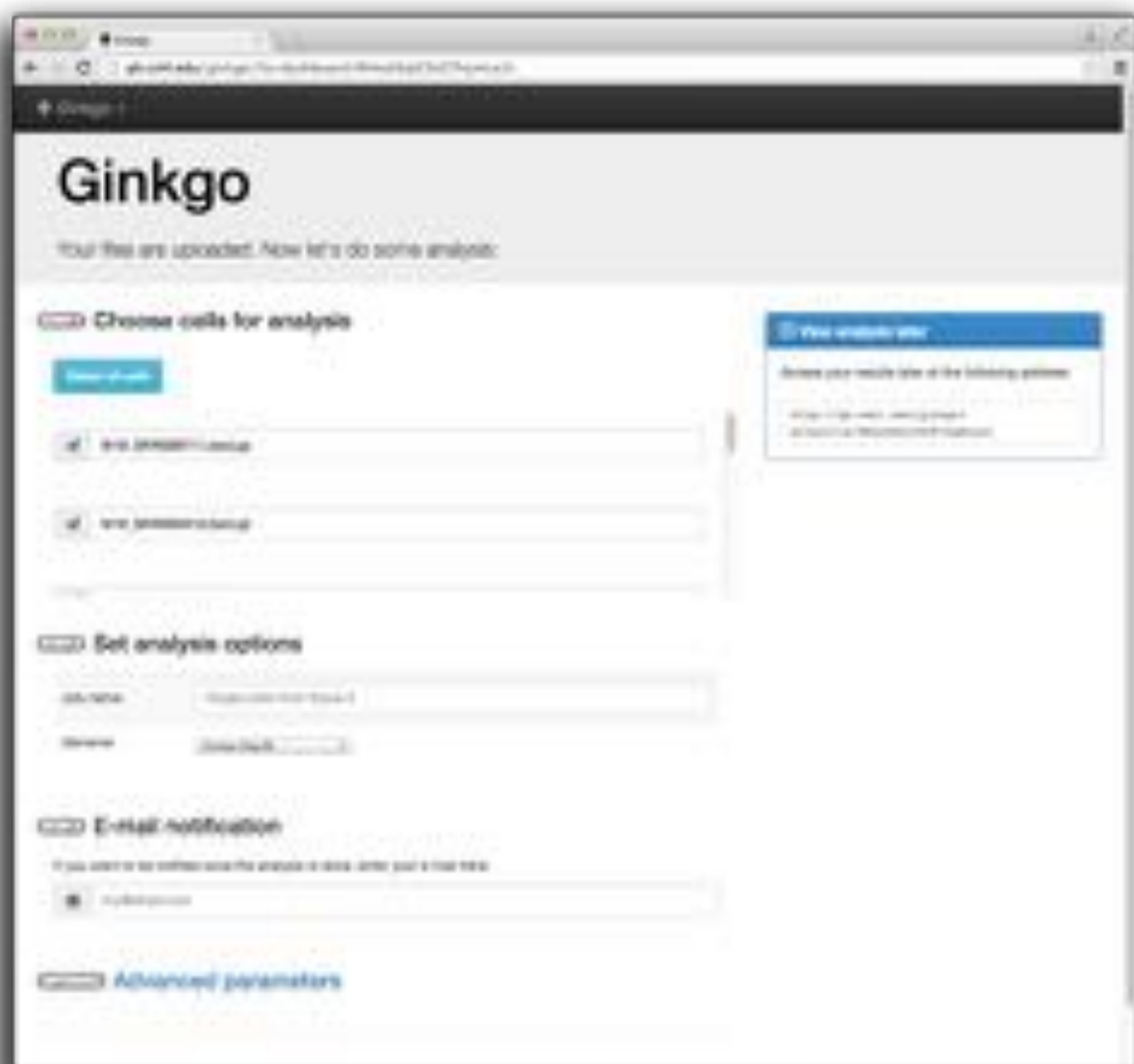


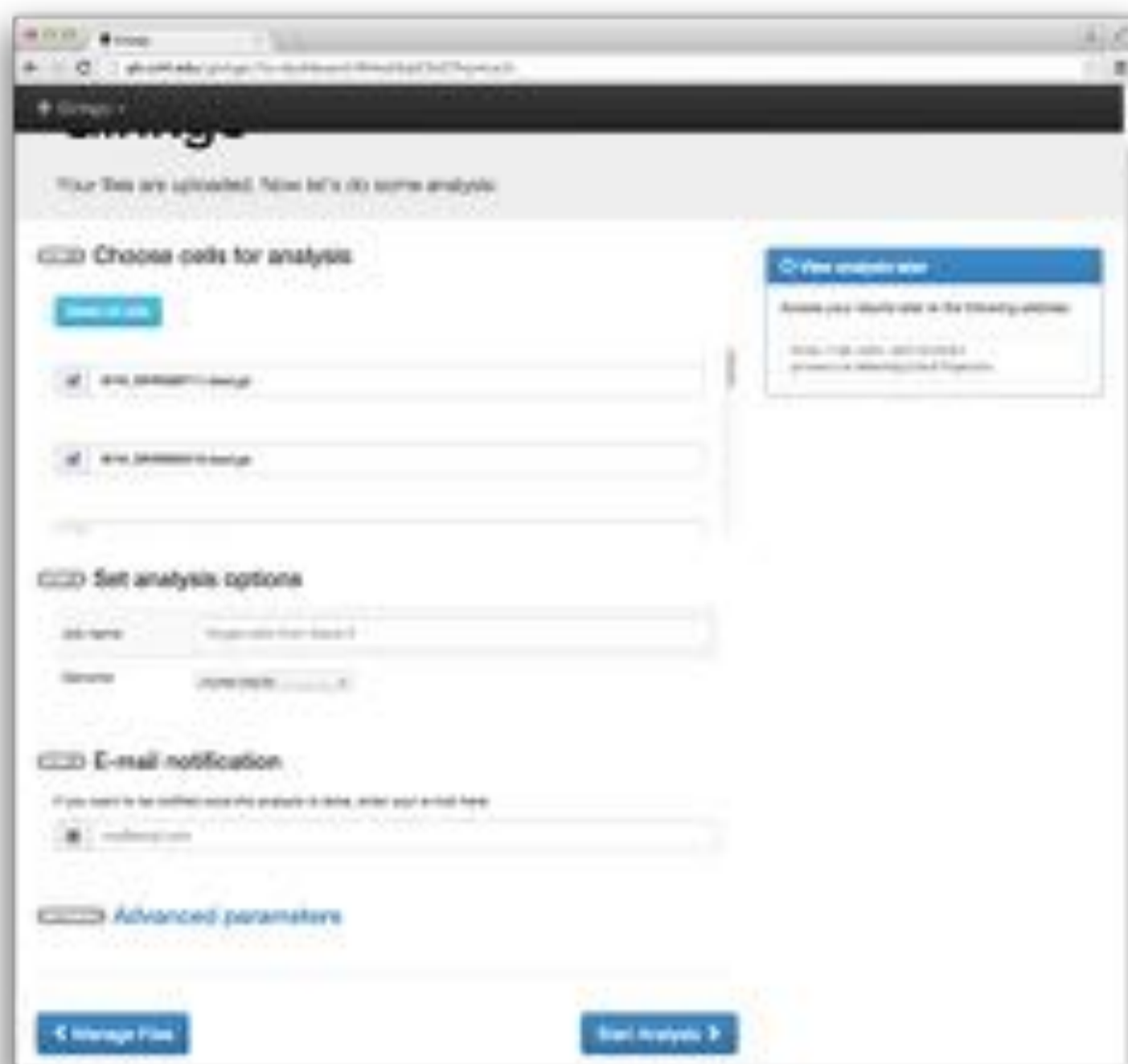
8 single cells from
breast tumor

Liver
Lesion

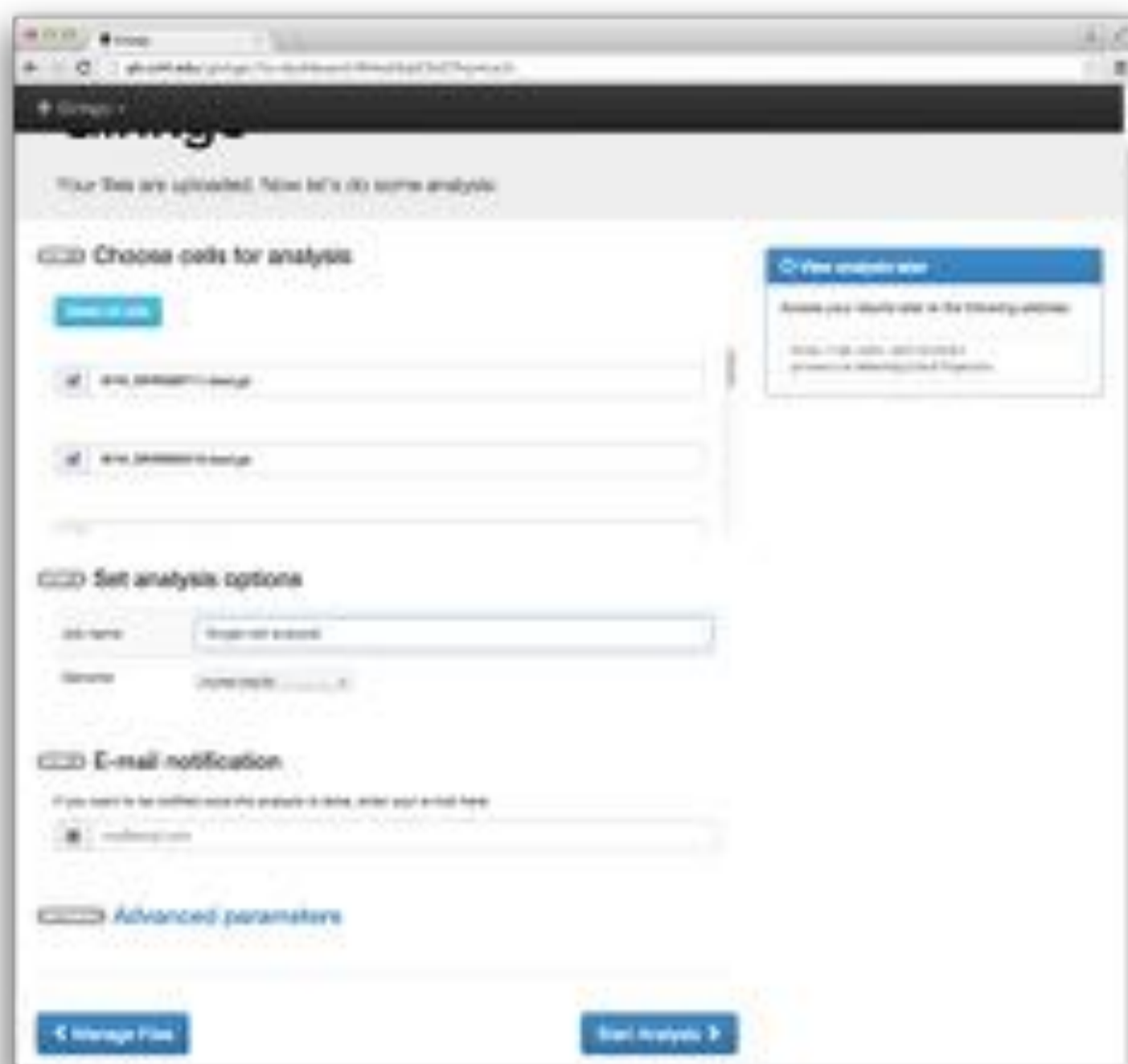


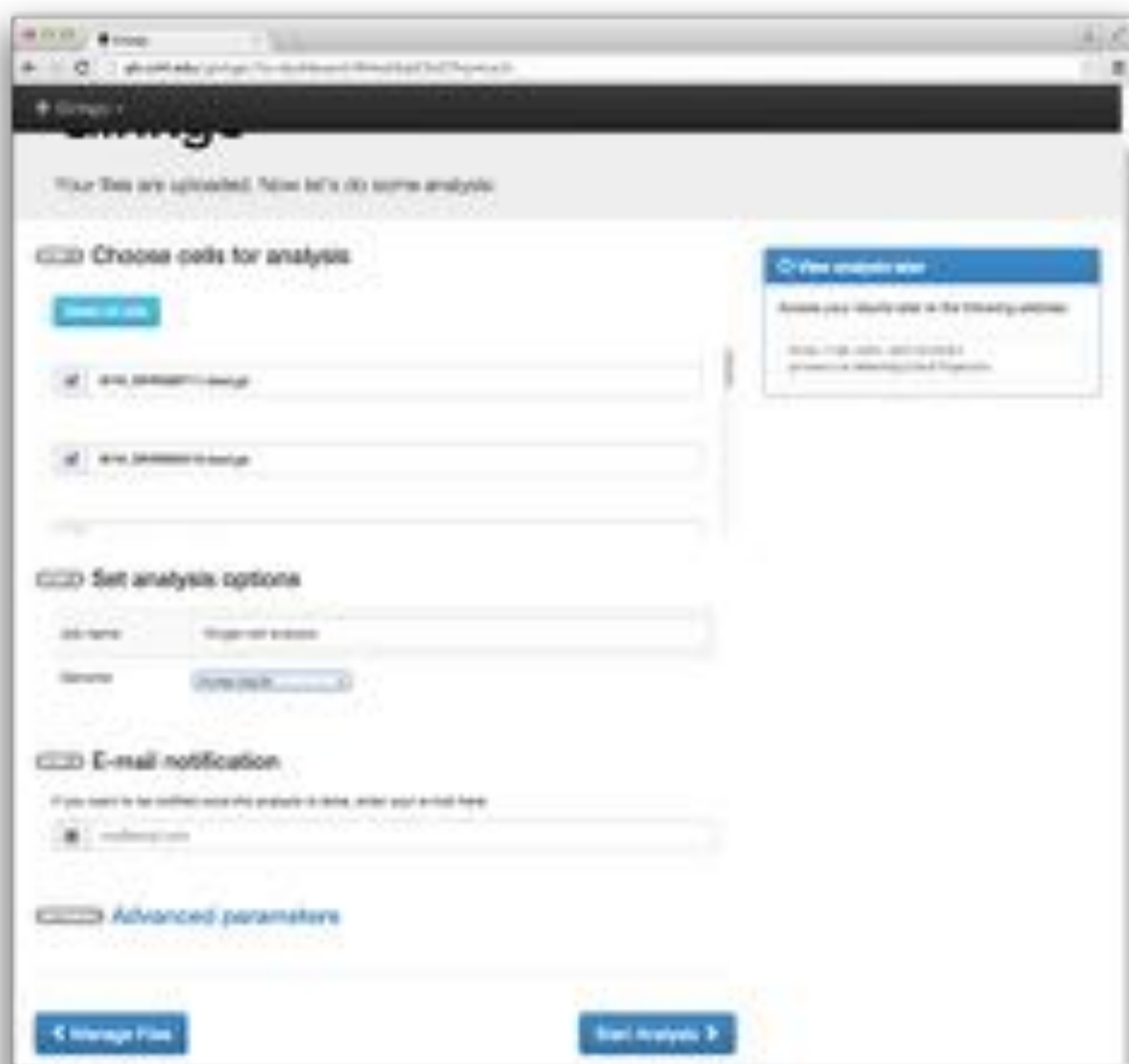
7 single cells from
liver tumor

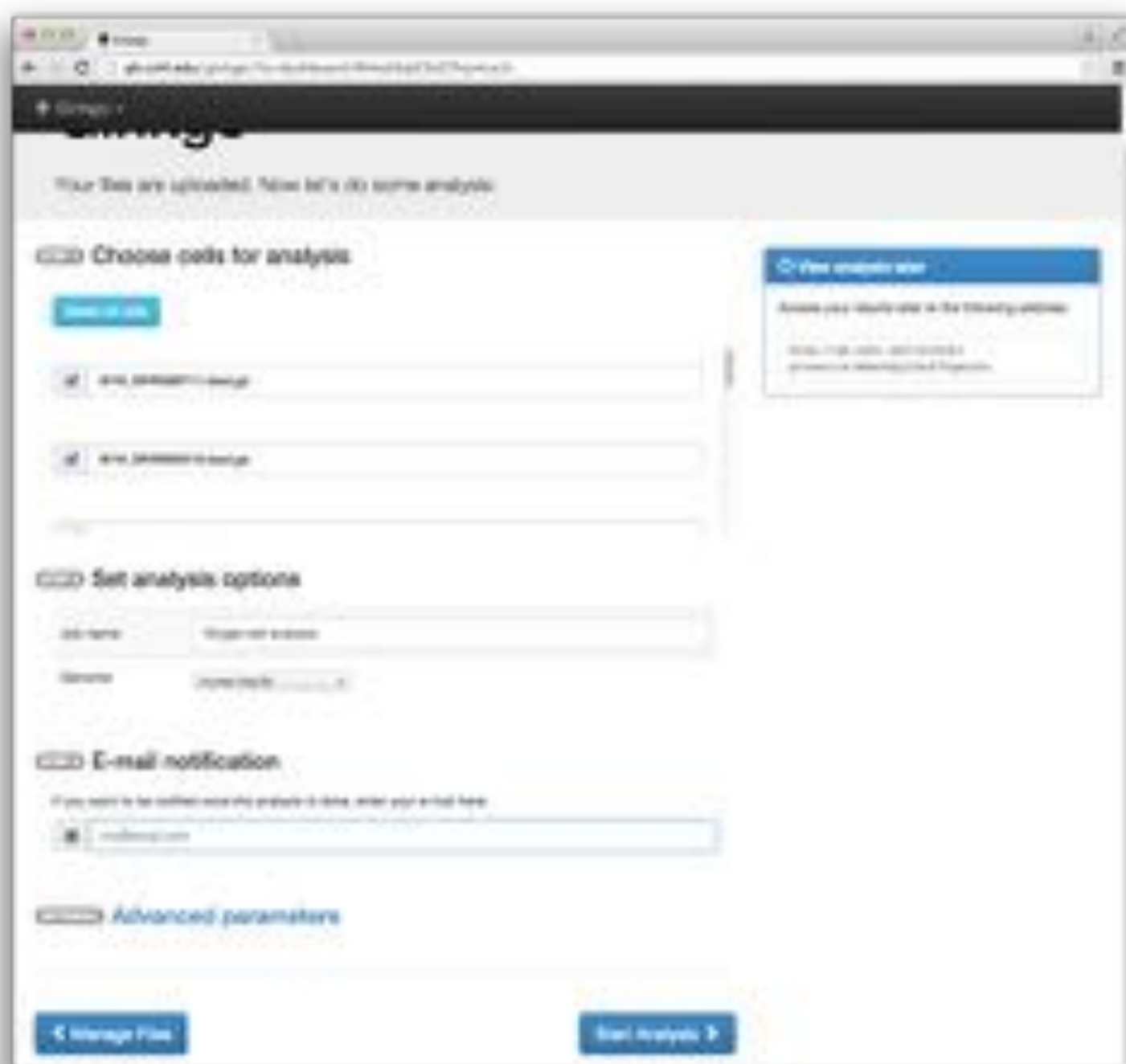


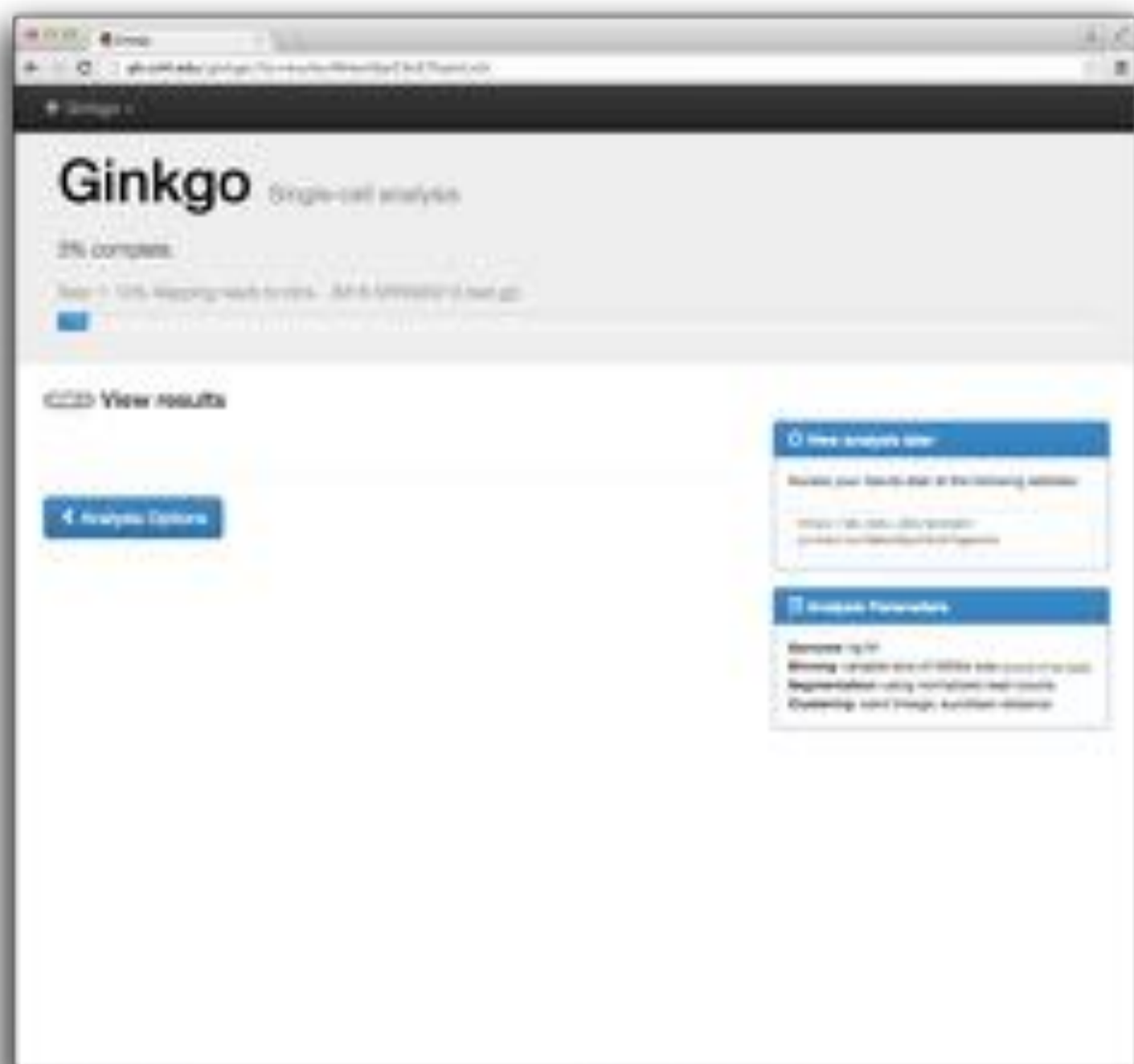


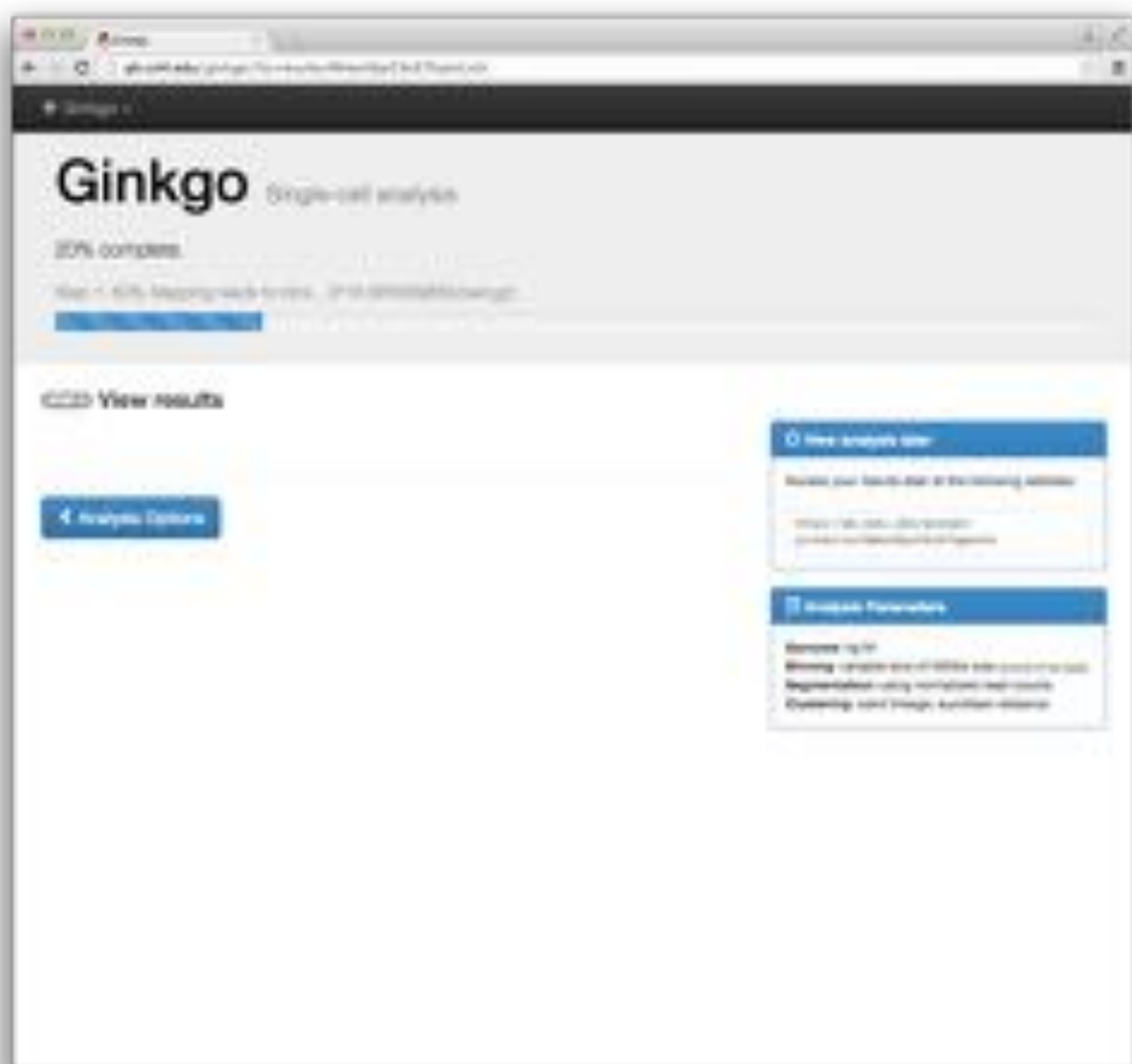


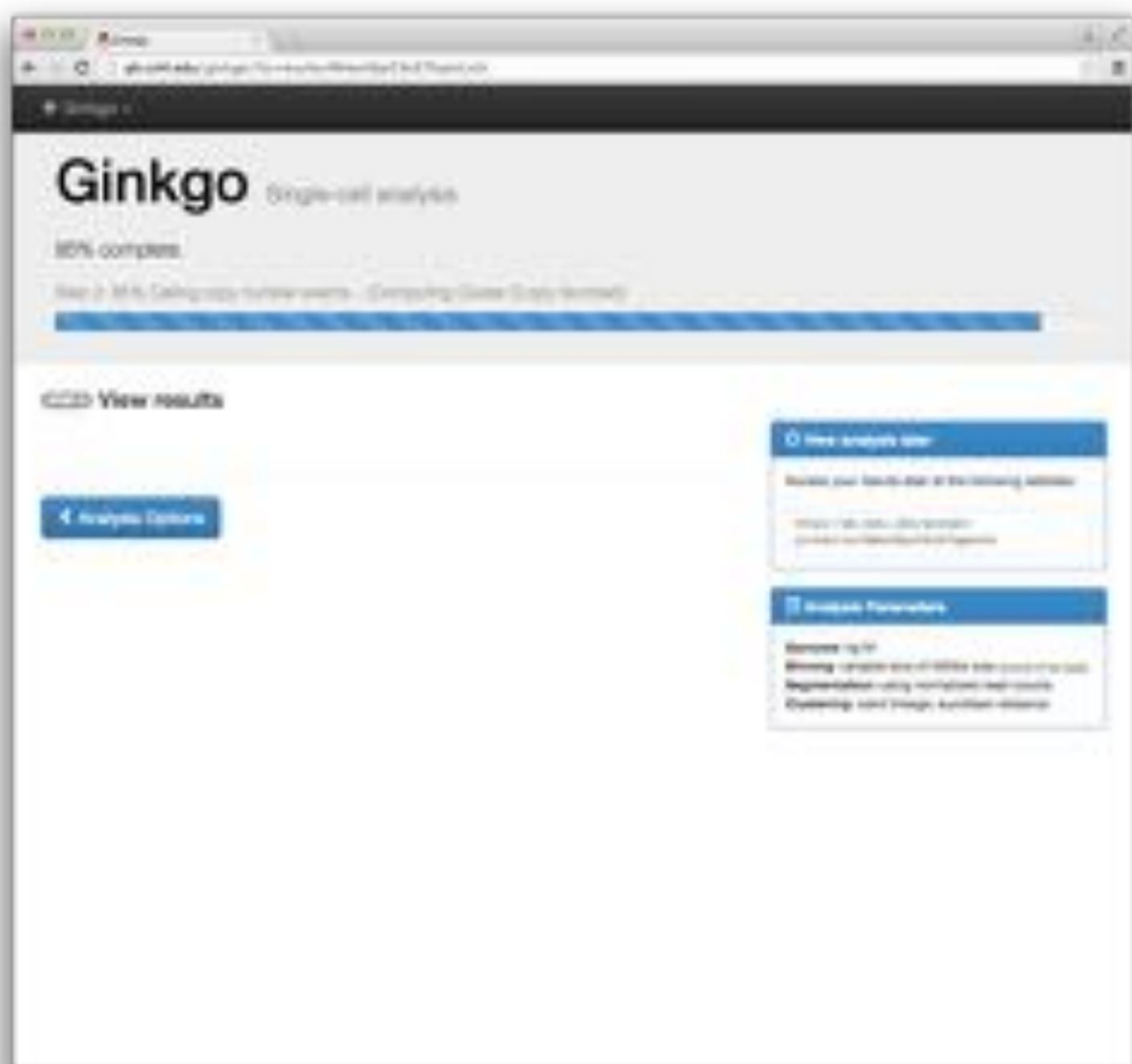










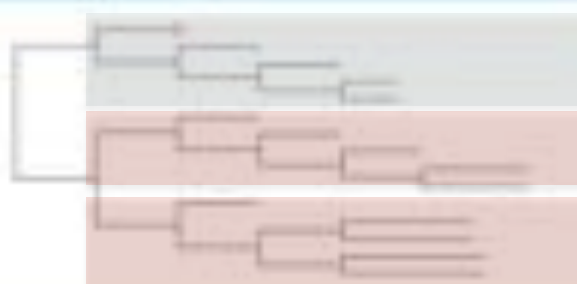


Ginkgo Original cell analysis

Analysing country scores

View results

• **Protein**



40-49	40-49
50-59	50-59
60-69	60-69
70-79	70-79
80-89	80-89
90-99	90-99
100-109	100-109
110-119	110-119
120-129	120-129
130-139	130-139
140-149	140-149
150-159	150-159
160-169	160-169
170-179	170-179
180-189	180-189
190-199	190-199
200-209	200-209
210-219	210-219
220-229	220-229
230-239	230-239
240-249	240-249
250-259	250-259
260-269	260-269
270-279	270-279
280-289	280-289
290-299	290-299
300-309	300-309
310-319	310-319
320-329	320-329
330-339	330-339
340-349	340-349
350-359	350-359
360-369	360-369
370-379	370-379
380-389	380-389
390-399	390-399
400-409	400-409
410-419	410-419
420-429	420-429
430-439	430-439
440-449	440-449
450-459	450-459
460-469	460-469
470-479	470-479
480-489	480-489
490-499	490-499
500-509	500-509
510-519	510-519
520-529	520-529
530-539	530-539
540-549	540-549
550-559	550-559
560-569	560-569
570-579	570-579
580-589	580-589
590-599	590-599
600-609	600-609
610-619	610-619
620-629	620-629
630-639	630-639
640-649	640-649
650-659	650-659
660-669	660-669
670-679	670-679
680-689	680-689
690-699	690-699
700-709	700-709
710-719	710-719
720-729	720-729
730-739	730-739
740-749	740-749
750-759	750-759
760-769	760-769
770-779	770-779
780-789	780-789
790-799	790-799
800-809	800-809
810-819	810-819
820-829	820-829
830-839	830-839
840-849	840-849
850-859	850-859
860-869	860-869
870-879	870-879
880-889	880-889
890-899	890-899
900-909	900-909
910-919	910-919
920-929	920-929
930-939	930-939
940-949	940-949
950-959	950-959
960-969	960-969
970-979	970-979
980-989	980-989
990-999	990-999
1000-1009	1000-1009
1010-1019	1010-1019
1020-1029	1020-1029
1030-1039	1030-1039
1040-1049	1040-1049
1050-1059	1050-1059
1060-1069	1060-1069
1070-1079	1070-1079
1080-1089	1080-1089
1090-1099	1090-1099
1100-1109	1100-1109
1110-1119	1110-1119
1120-1129	1120-1129
1130-1139	1130-1139
1140-1149	1140-1149
1150-1159	1150-1159
1160-1169	1160-1169
1170-1179	1170-1179
1180-1189	1180-1189
1190-1199	1190-1199
1200-1209	1200-1209
1210-1219	1210-1219
1220-1229	1220-1229
1230-1239	1230-1239
1240-1249	1240-1249
1250-1259	1

[View comments on this](#)

Answers will vary, but should include the following:

Table 1

Abstract

Keywords: *language, social interaction, communication, pragmatics, early communication, autism*

• [Sign in](#)

(c) ~~Document was not reviewed for copyright~~

Keep number of items at least 100.

www.merck.com

Downloaded by [University of Cambridge] at 11:53 11 September 2015

© 2004 Blackwell Publishing Ltd *Journal of Internal Medicine* 255: 105–112

It is important to note that possible outliers could not be identified in this analysis. If a researcher wanted to identify outliers in this dataset, they could use the following formula:

Image: <http://www.royalcanin.com> (royalcanin.com) is a website for pet owners. The website is a good example of a website that is easy to use and provides a lot of information.

[Downloaded from](#) [http://ajphaphysoc.org/](#) on November 11, 2015

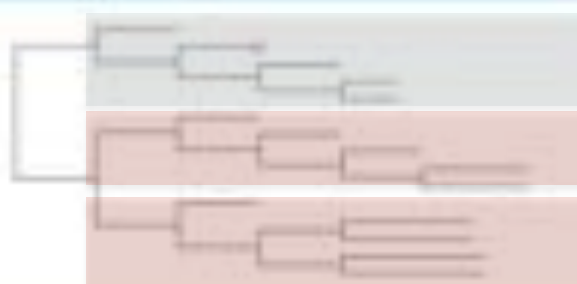
[illegible]

Ginkgo Single-cell analysis

Analysing country scores

View results

• **Protein**

[illegible]

[View comments on this](#)

described in a recently open public bidding process.

100

100%

Keywords: *biology, ecology, and of others and ...*
Supplemental ...
Quarterly ...

• **Answer:** *See above.*

Cell	QW-Poly	W-Route	Mean total count	Mean count variance	Index of dispersion
Cell 10000001		10,011,000	1,000,000	1,000,000	0.00
Cell 10000002		11,000,000	1,100,000	1,100,000	0.00
Cell 10000003		11,000,000	1,100,000	1,100,000	0.00
Cell 10000004		11,000,000	1,100,000	1,100,000	0.00
Cell 10000005		11,000,000	1,100,000	1,100,000	0.00
Cell 10000006		11,000,000	1,100,000	1,100,000	0.00

■ **Group discussion:**

(c) *Documentary evidence* (see paragraph 10.1.1)

© 2004 Blackwell Publishing Ltd *Journal of Internal Medicine* 255: 105–112

www.elsevier.com/locate/jmb

Downloaded by [University of Cambridge] at 18:00 11 September 2015

© 2004 Blackwell Publishing Ltd *Journal of Internal Medicine* 255: 105–112

REPRODUCIBLE: Items that produce identical results that are interpreted in the same manner. 1 = completely reproducible; 2 = 3 = moderately & very reproducible; 4 = 5 = not reproducible.

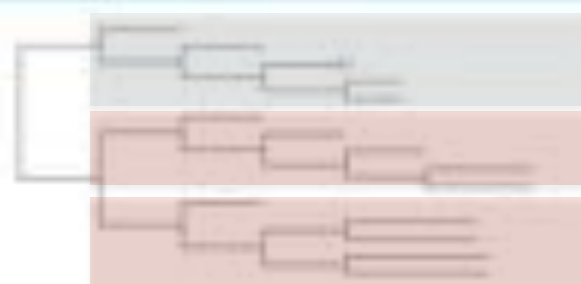
Source: *Statistical Abstract of the United States*, 1997, table 1201. <http://www.census.gov/states/ny>.

Ginkgo Original cell analysis

Analysing country scores

View results

• **Protein**



[View comments on this](#)

Accepted 10 June 2005

... ..

[illegible]

• **Answer:** *See above.*

Cell	QW Profile	W Bandwidth	Mean read current	Peak current (nA)	Index of Dispersion
Cell 1, channel 1		10.011.000	0.000.000	1.000.000	0.00
Cell 1, channel 2		11.000.000	0.000.000	1.000.000	0.00
Cell 1, channel 3		11.000.000	0.000.000	1.000.000	0.00
Cell 1, channel 4		11.000.000	0.000.000	1.000.000	0.00
Cell 1, channel 5		11.000.000	0.000.000	1.000.000	0.00
Cell 1, channel 6		11.000.000	0.000.000	1.000.000	0.00

■ **Group discussion:**

(a) **Documented case** (see comment on 10/11/2014)

These numbers should be left blank.

transboundary systems and networks

Downloaded by [University of Cambridge] at 11:52 11 September 2015

Received 10 May 2006; accepted 10 May 2006

[illegible]

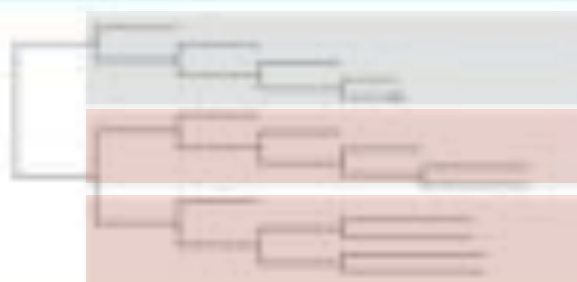
Source: National Highway Traffic Safety Administration, Bureau of Motor Vehicle Safety, Washington, DC, 2000.

Ginkgo Single-cell analysis

Analysing country scores

View results

• **Protein**



The 'New' button is highlighted in the 'New' dialog box.

[View comments and join the discussion](#)

Answers are provided at the end of each chapter.

...and the ...

100

...and the...
...and the...
...and the...

• **Answer:** *See above.*

Cell	Open Facility	W. Results	Mean total cost	Actual capacity utilization	Index of Dispersion
1) K12_200000711		10,817,148	2,868,145	1,625.36	0.58
2) K12_200000712		11,569,868	2,925,497	1,517.29	0.58
3) K12_200000714		11,363,130	2,918,138	1,520.94	0.47
4) K12_200000716		1,508,037	1,613,105	166.23	0.48
5) K12_200000718		4,814,240	1,736,105	278.08	0.48

■ **Group discussion:**

(c) **Documented case** (see comment on last cell)

These numbers provide an excellent guide.

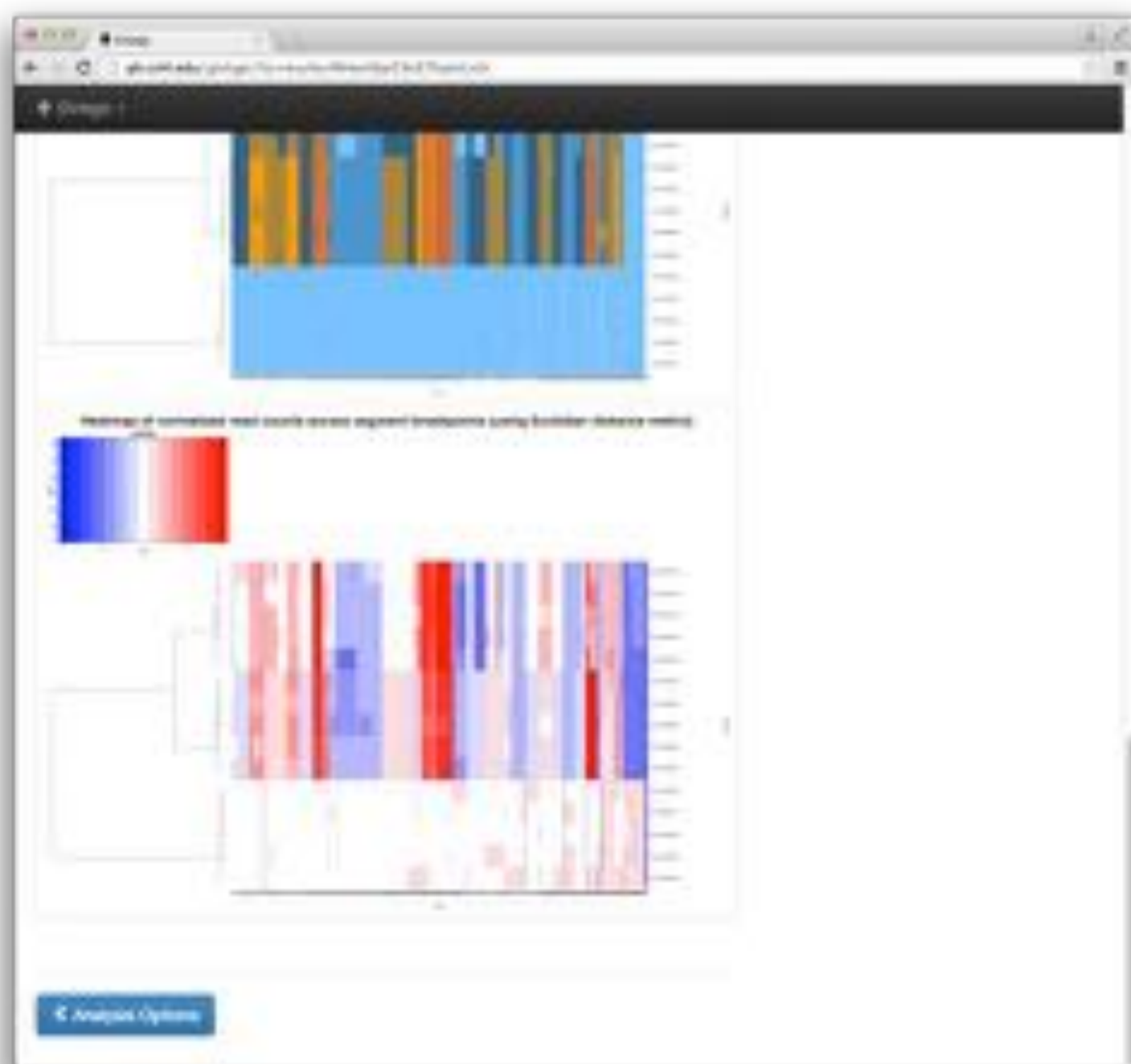
transboundary systems and networks

Downloaded by [University of Cambridge] at 11:53 11 September 2015

Received 10 November 2003; accepted 10 November 2003

© 2004 Blackwell Publishing Ltd, *Journal of Internal Medicine* 255: 101–108

Source: National Highway Traffic Safety Administration, Bureau of Motor Vehicle Safety, Washington, DC, 2000.



Ginkgo Single-cell analysis

Analysis completed

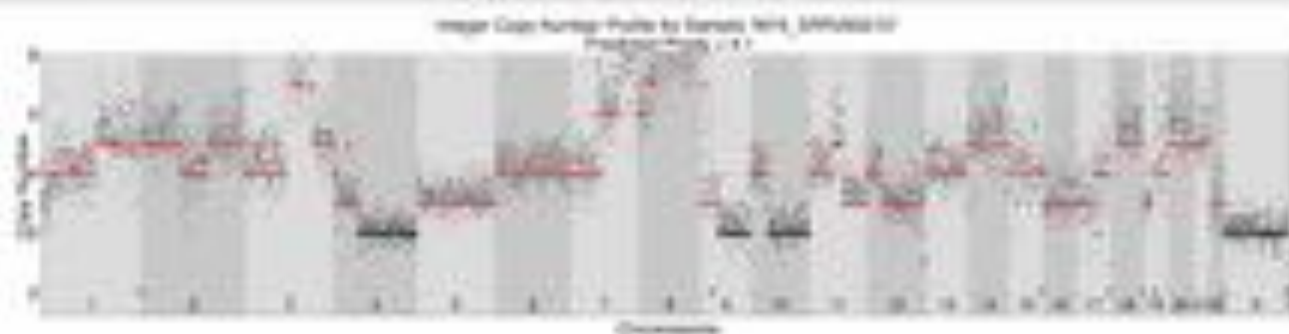
Viewing call M14_20000210

4. March 1999

© 2004 Blackwell Publishing Ltd *Journal of Internal Medicine* 255: 105–112

[View Paper on Scribd](#)

■ **Classroom Connections**



Ginkgo Single-cell analysis

Analysis comments

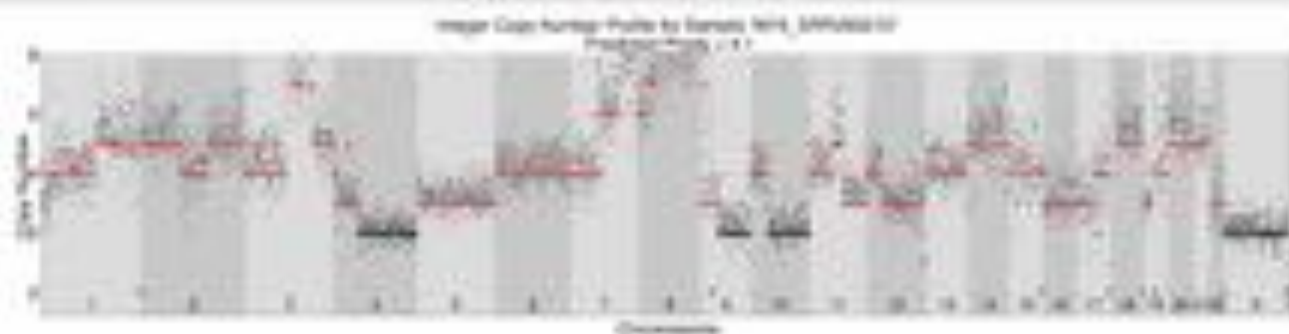
Viewing call M14_00000210

4. [May 15, 2016](#)

© 2004 Blackwell Publishing Ltd *Journal of Internal Medicine* 255: 105–112

[View Paper on Scribd](#)

■ **Classroom Connections**



Ginkgo Single-cell analysis

Analysis components

Viewing call M14_20000210

4. March 1999

► *Journal of Management Education*

Downloaded from <http://ajphaphysocpharm.sagepub.com/>

[illegible]

© 2005 Blackwell Publishing Ltd *Journal of Internal Medicine* 258: 259–266

[illegible]

Downloaded At: 11:53 11 September 2009

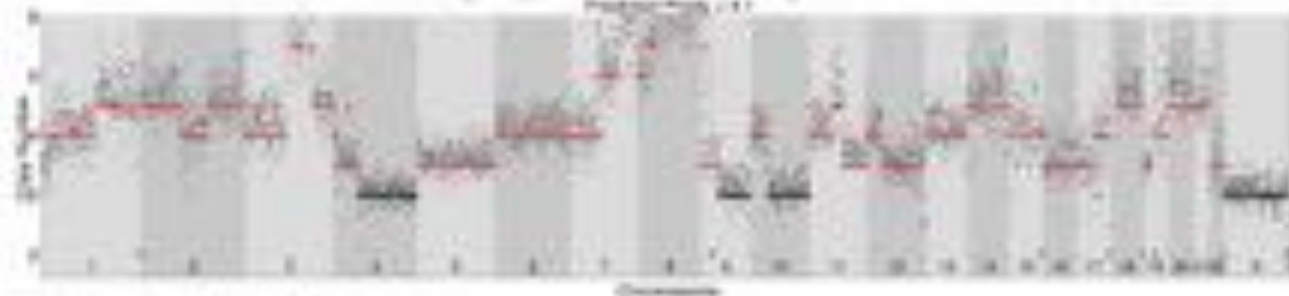
Figure 1

© 2000 Blackwell Science Ltd

■ **Classroom Connections**

Source: U.S. Census Bureau, *Current Population Reports*, 1990.

1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 26





Ginkgo Single-cell analysis

Analysis completed

Viewing call M14_00000210

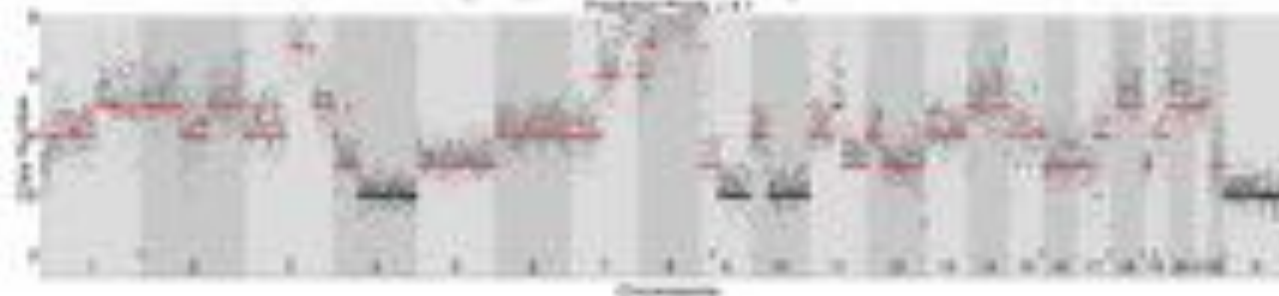
4 March 2016

© 2004 Blackwell Publishing Ltd *Journal of Internal Medicine* 255: 103–110

[View Paper on Scribd](#)

■ **Classroom Connections**

Image Courtesy: Photo by Samson Neri, *Shutterstock*
(freemovephoto.com)



Ginkgo Single-cell analysis

Analysis completed

Viewing call ME4_00000210

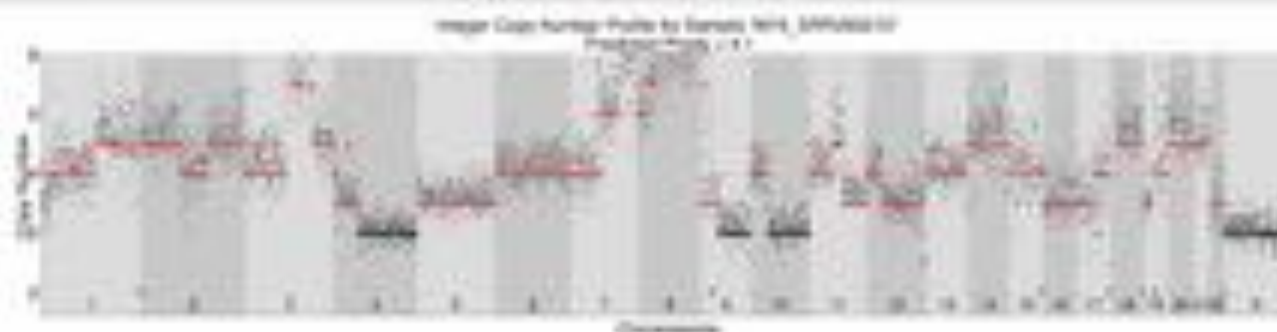
4 March 2016

► *Journal of Management Education*

Free download of PDF books
<http://www.pdfbooks.org>



■ **Classroom Connections**



Ginkgo Single-cell analysis

Analysis completed

Viewing call ME4_00000210

4 March 2016

© 2004 Blackwell Publishing Ltd *Journal of Internal Medicine* 255: 103–110

[View Paper on Scribd](#)

© 2004 Blackwell Publishing Ltd

© 2005 Blackwell Publishing Ltd *Journal of Internal Medicine* 258: 105–112

Abstract

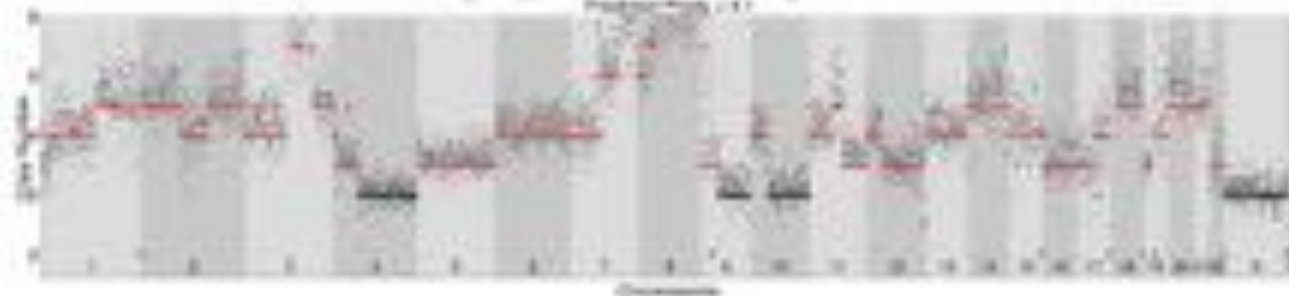
Strategic analysis

100

© 2006 Blackwell Publishing Ltd *Journal of Internal Medicine* 260: 395–403

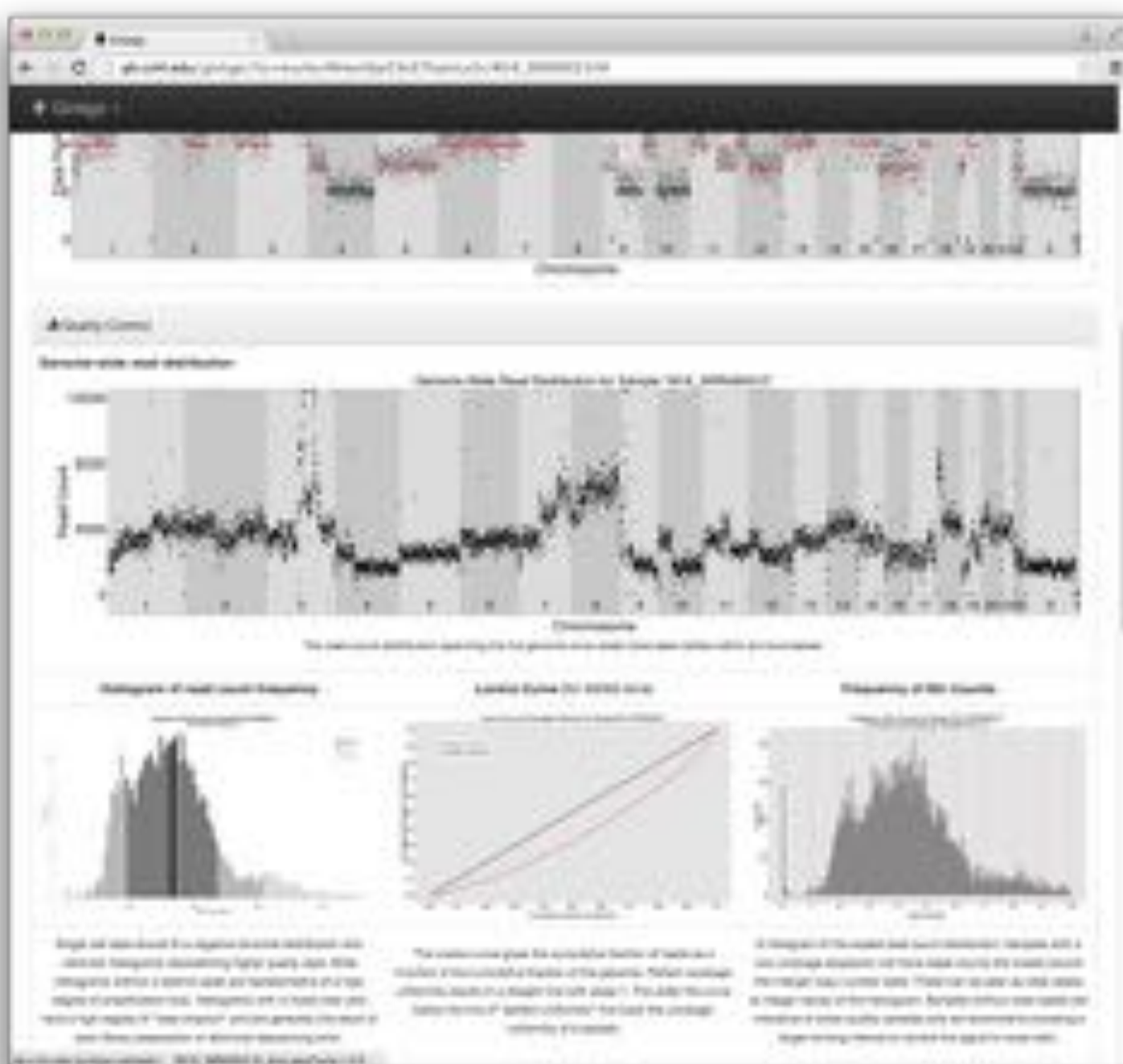
■ **Classroom Connections**

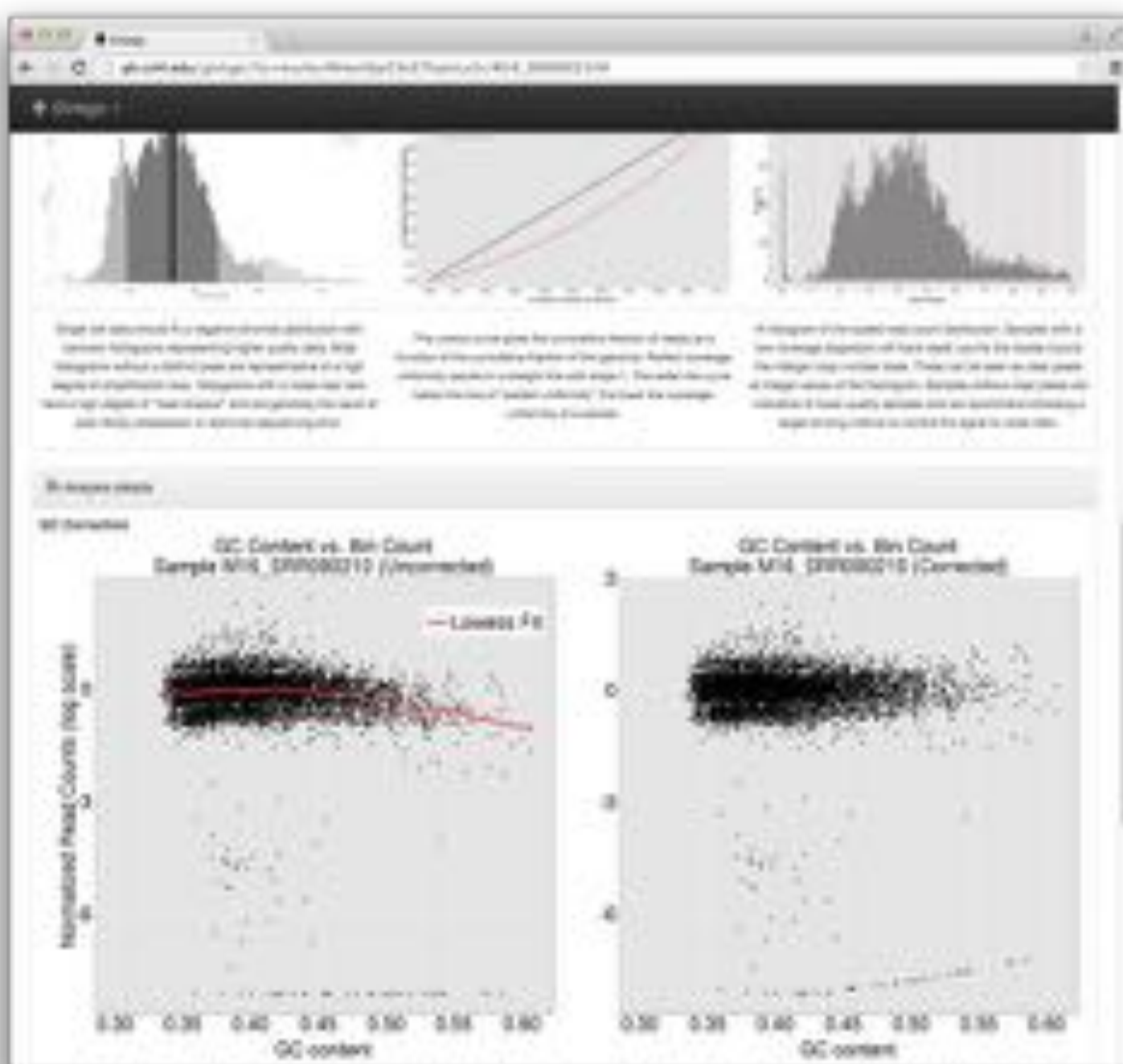
Image Courtesy: Photo by Samson Neri, *Shutterstock*
(freemovephoto.com)

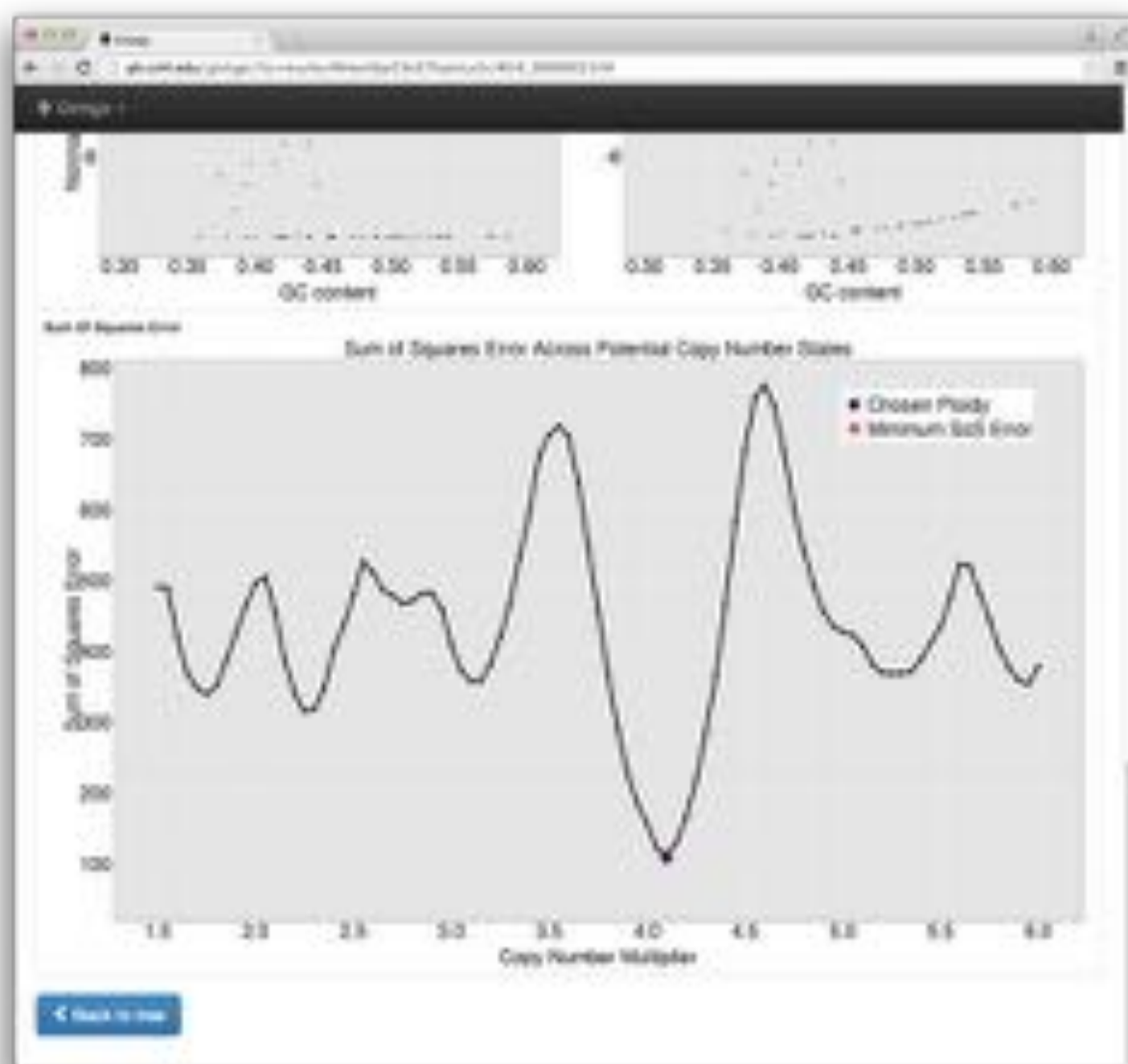










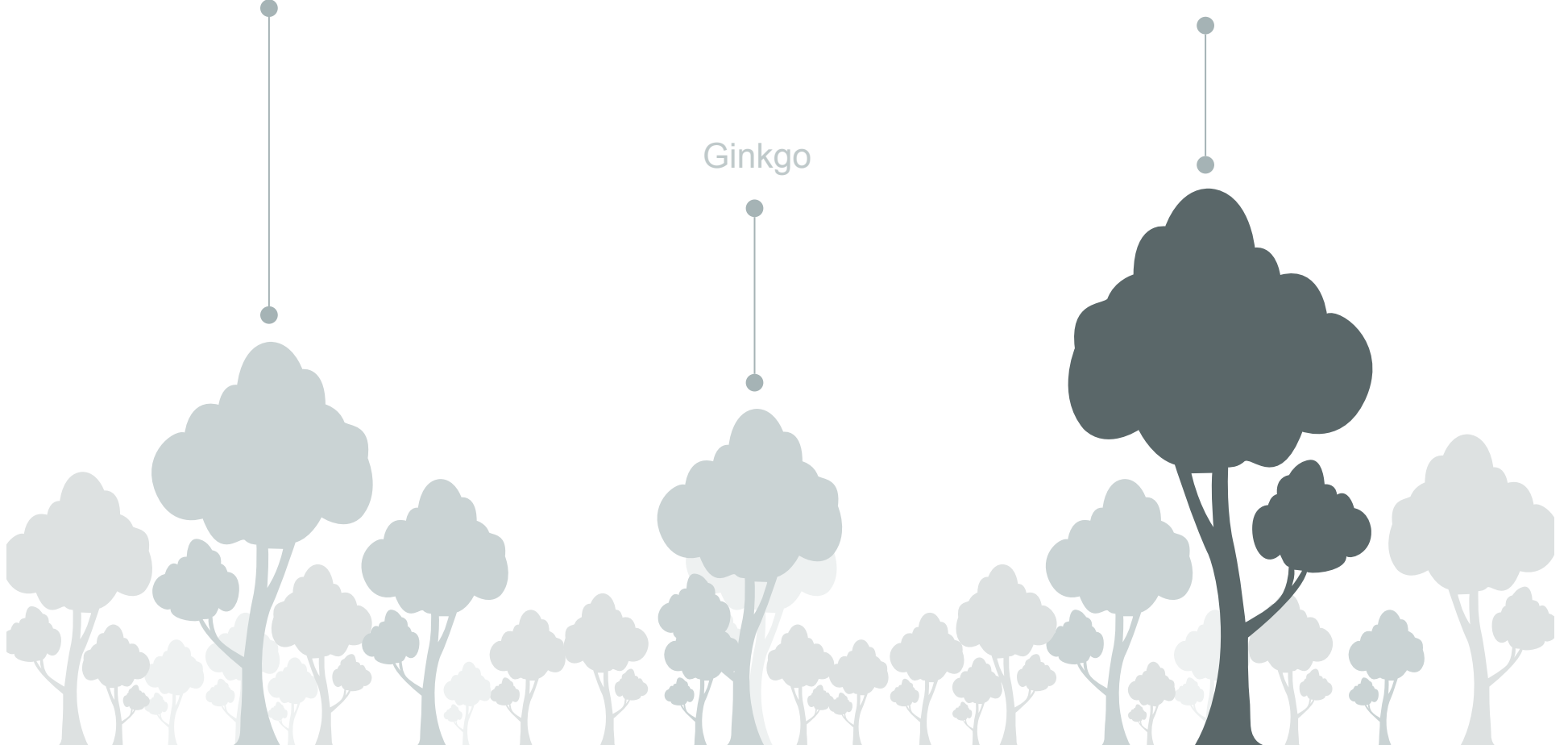


Outline

Introduction

Comparison of
WGA methods

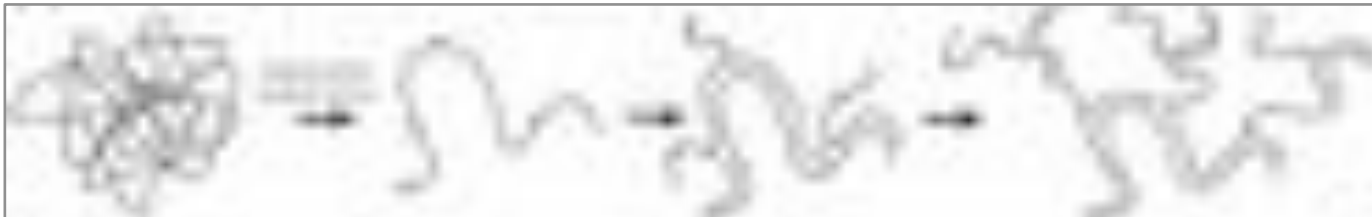
Ginkgo



Whole Genome Amplification (WGA) methods



DOP-PCR (Degenerate Oligonucleotide Primed PCR)



MDA (Multiple Displacement Amplification)



MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

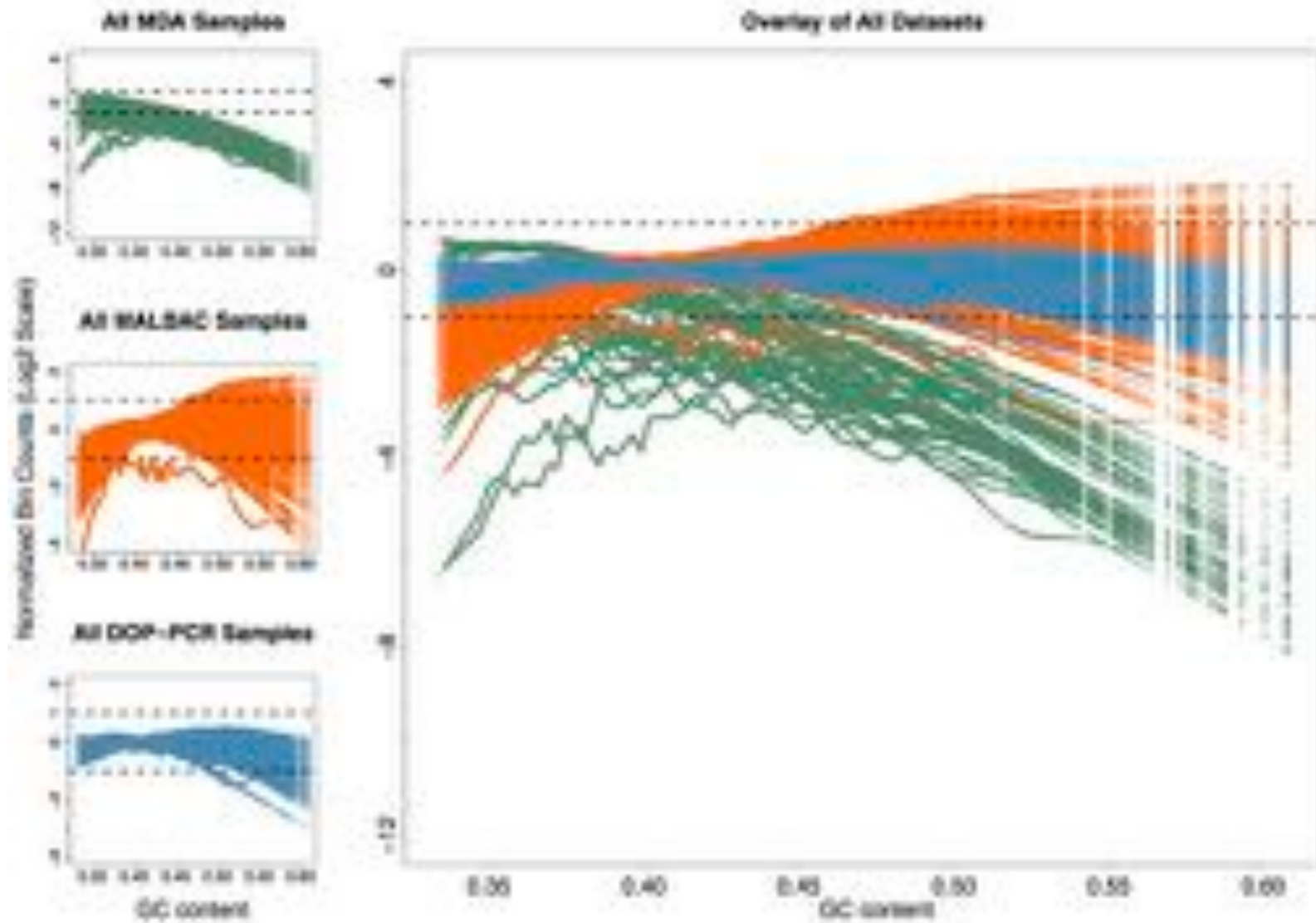
Comparison of WGA methods

Paper	WGA Method	Tissue
Navin et al., 2011	DOP-PCR	Breast (T10)
Navin et al., 2011	DOP-PCR	Breast (T16P/M)
McConnell et al., 2013	DOP-PCR	Neuron
Lu et al., 2012	MALBAC	Sperm
Ni et al., 2013	MALBAC	Lung
Hou et al., 2013	MALBAC	Oocyte
Kirkness et al., 2013	MDA	Sperm
Wang et al., 2012	MDA	Sperm
Evrny et al., 2012	MDA	Neuron

Explore the effects of WGA method on data quality:

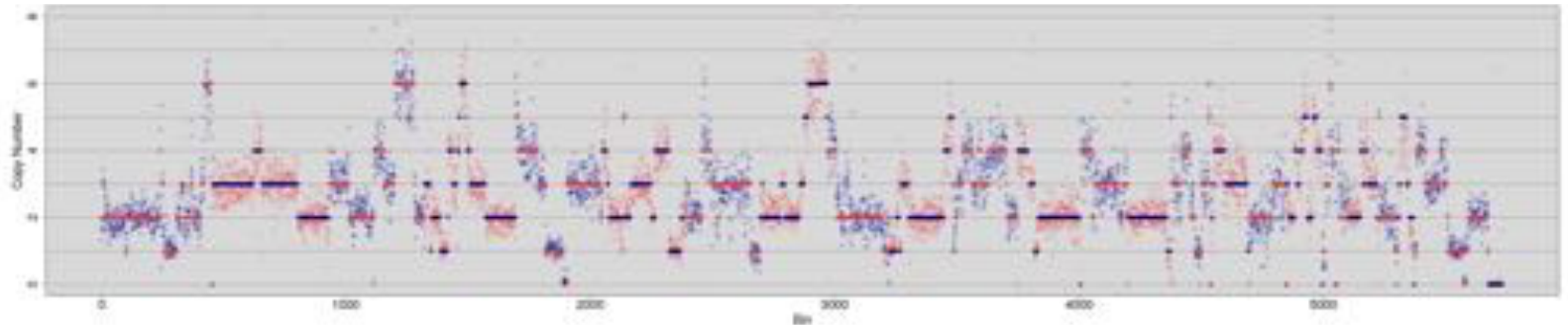
- 1) GC bias
- 2) Coverage dispersion

GC Bias

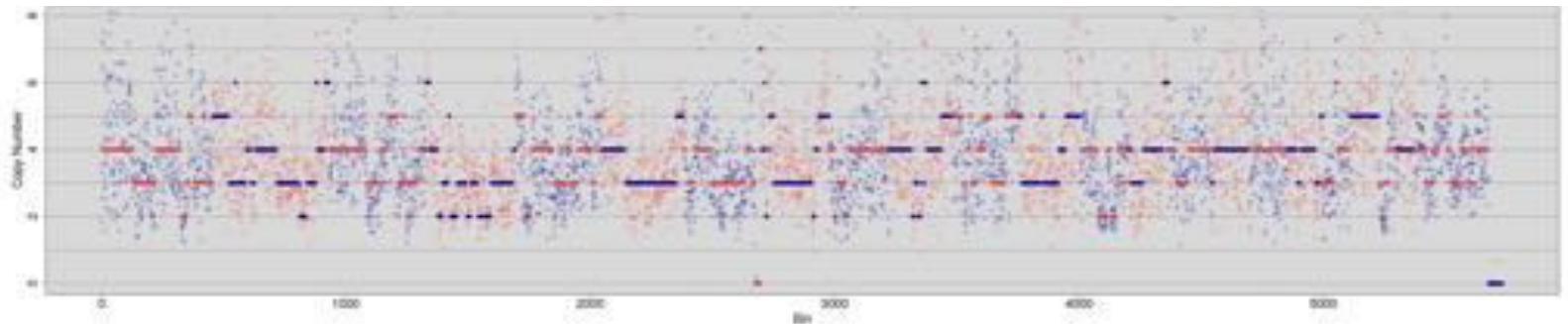


Coverage Dispersion

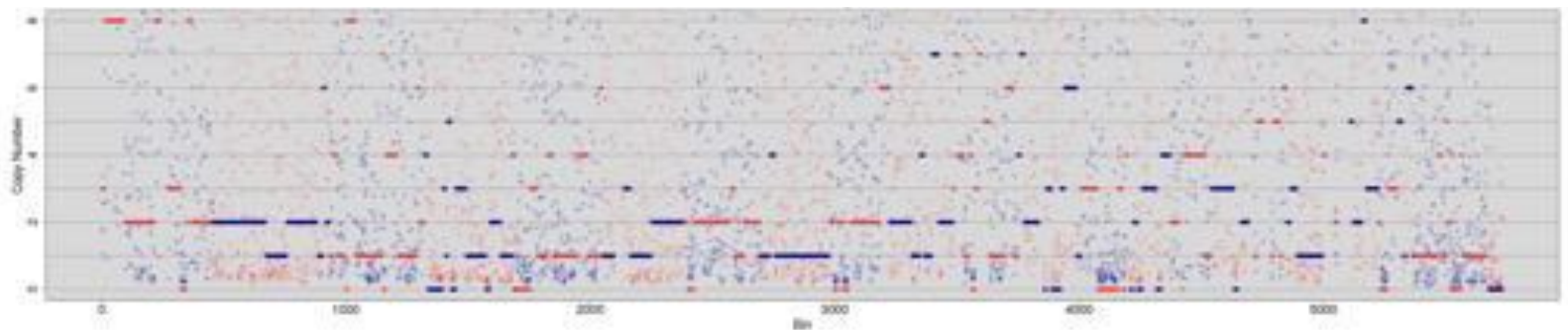
DOP-PCR



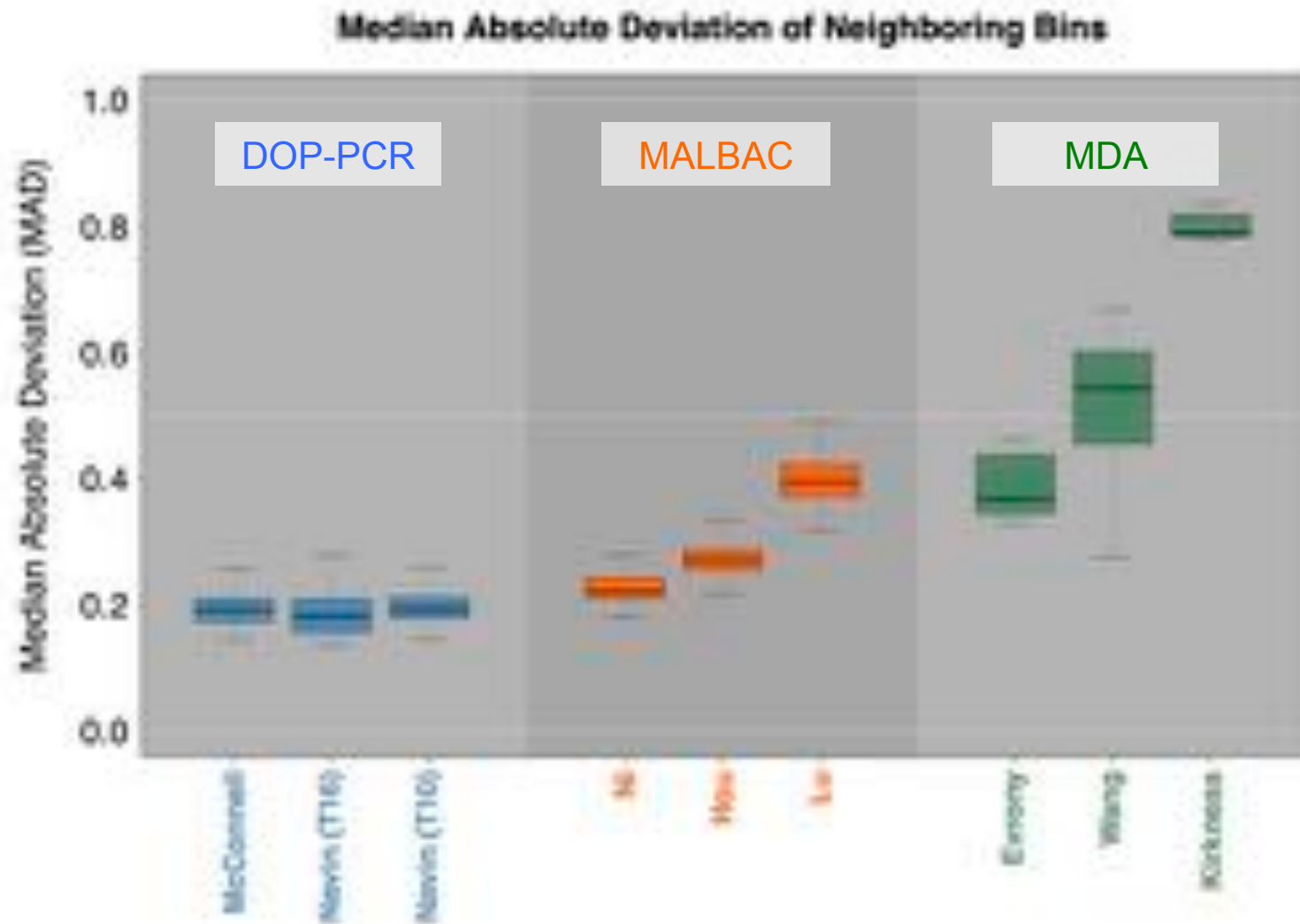
MALBAC



MDA



Coverage Dispersion



Summary

- Ginkgo is a platform for single-cell CNV analysis and visualization
- For copy-number analysis, we recommend DOP-PCR
- Check out Ginkgo and give us feedback
 - qb.cshl.edu/ginkgo
 - <http://qb.cshl.edu/ginkgo/workshop/fog.pdf>
 - *Garvin and Aboukhalil et al., Nature Methods, 2015*

> 1K
USERS

21,500
PAGEVIEWS

10 MIN
TIME SPENT

Thanks

Ginkgo Team

Rob Aboukhalil

Jude Kendall

Timour Baslan

Jim Hicks

Gurinder S. Atwal

Michael Wigler

Michael C. Schatz

qb.cshl.edu/ginkgo

