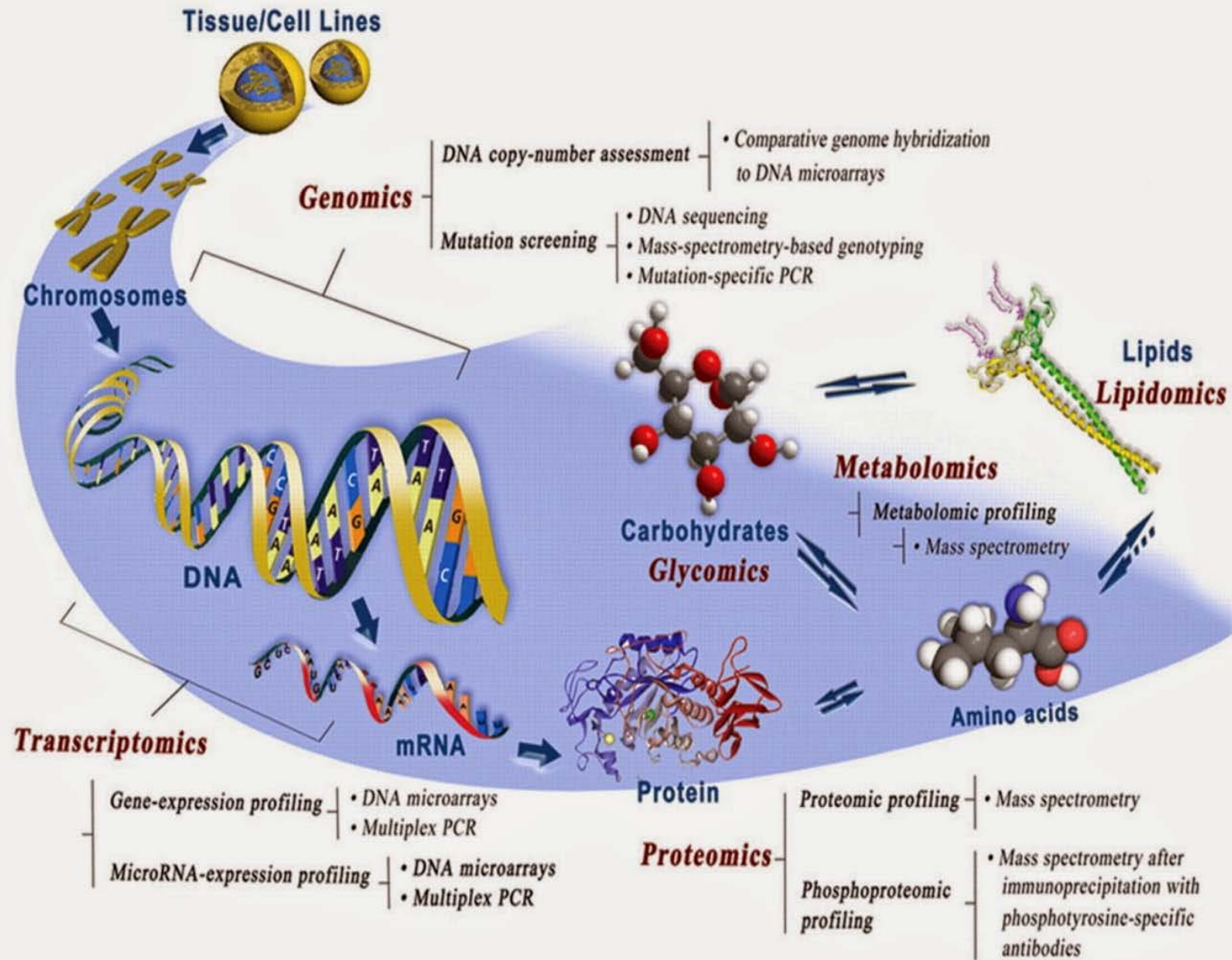


Omics technologies: **genomics, transcriptomics, metabolomics,** **databases, personalized medicine and big data**

Vendula Pospíchalová, PhD
(pospich@sci.muni.cz)
Department of Experimental Biology
Animal Physiology and Immunology

Bi5599 Applied Biochemistry and Cell Biology Methods
2021-11-08



Schematic representation of omics technologies, their corresponding analysis targets, and assessment methods. Taken from Wu RD et al. JDR 2011; 90:561-572.

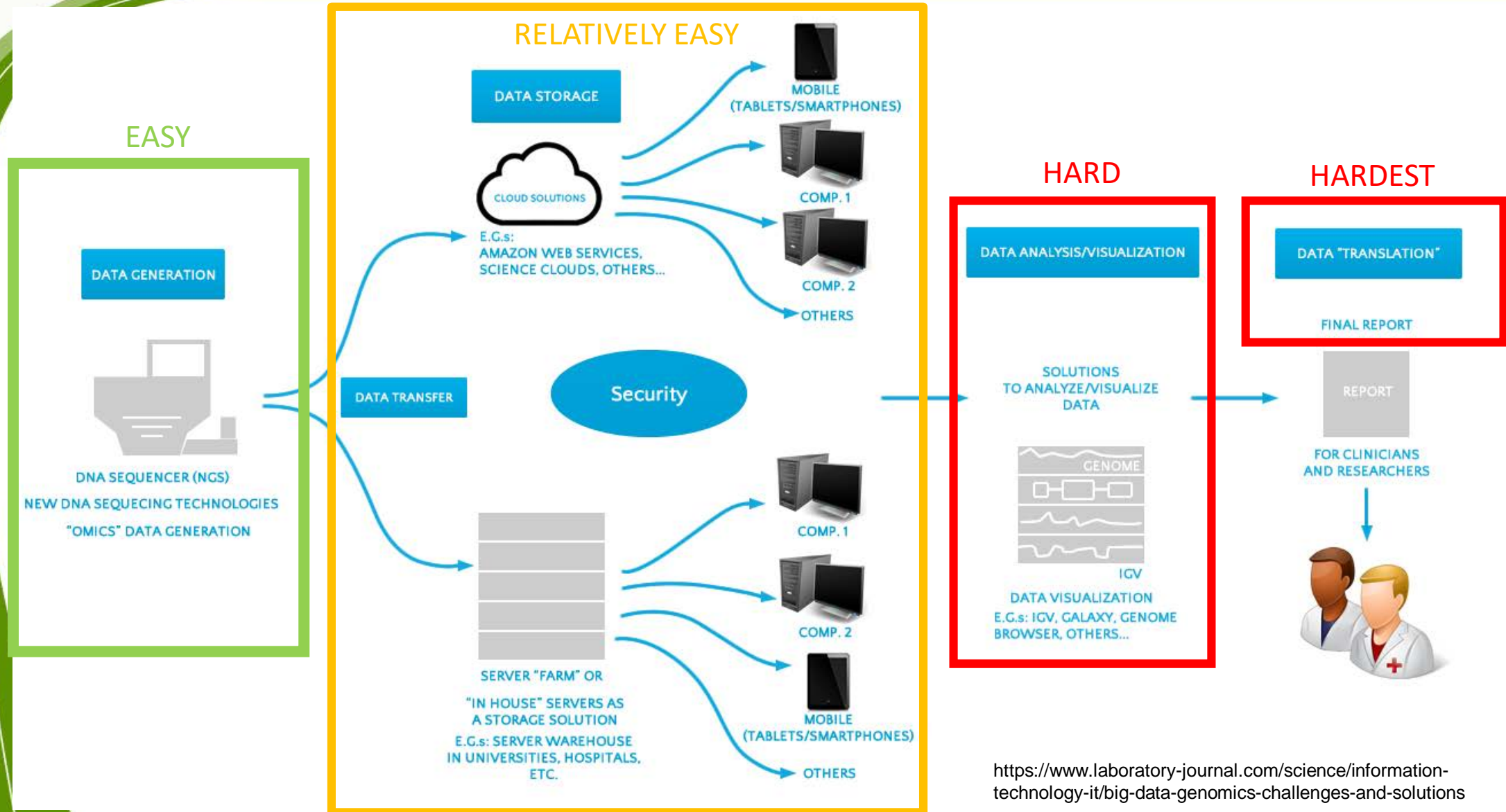
Contents

1. Introduction: what are –omics technologies + history
2. How does big data look and how to approach it
3. From –omics technologies to biomarkers and personalized medicine
4. Genomics: genomes vs exomes vs genotypes + DTC service
5. Cancer databases: COSMIC, TCGA and others
6. Transcriptomics: microarrays vs. RNA sequencing
7. Gene set analysis
8. Metabolomics
9. Cutting edge: single cell –omics and single cell multi-omics
10. Summary and take home messages

What are „-omics“ technologies

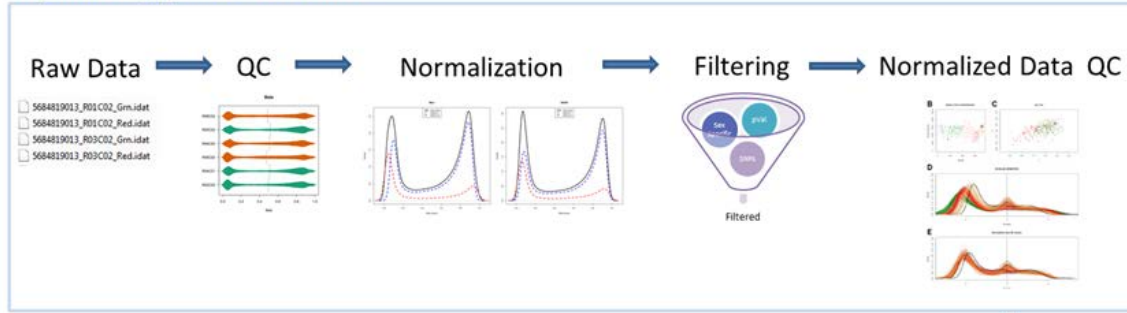
- **Omics** refers to a field of study in biology ending in **-omics**, such as genomics, proteomics or metabolomics
- The related suffix **-ome** is used to address the objects of study of such fields, such as the genome, proteome or metabolome
- **-ome** = many/collectivity or whole/all/complete in Greek
- **-omics** = study of large sets of biomolecules
- **High-throughput experimental technologies** characterized by **automation, miniaturized assays** and **large-scale data analysis**
- Analytic part of the experiment is usually much longer than the experiment itself – bioinformatics skills needed
- Raw data is the „gem“ but usually is in user unfriendly format
- **Interpreting** functional consequences of millions of discovered events is one of **the biggest challenges**

Big -omics data challenges

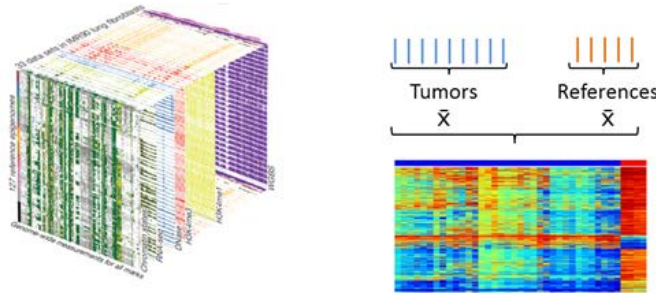


Only skilled bioinformaticians can process raw data

optimized pipeline with filters



Functional Genomics ← Supervised Analysis ← Unsupervised Analysis



Integrated analyses of –omics studies are only possible in large consortia (e.g. TCGA). Subsequently, the authorlist of such articles can more than 2 pages long with substantial part of the authors being bioinformaticians. Among the reviewers, bioinformaticians are also necessary, etc.

ARTICLE

Integrated genomic analyses of ovarian carcinoma

The Cancer Genome Atlas Research Network*

- doi:10.1038/nature10166
30. Dresman, H. K. et al. An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *J. Clin. Oncol.* **25**, 517–525 (2007).
31. Creighton, C. J. et al. Insulin-like growth factor 1 activates gene transcription programs strongly associated with poor breast cancer prognosis. *J. Clin. Oncol.* **26**, 4070–4080 (2008).
32. Kashiba Prasad, T. S. et al. Human Protein Reference Database—2009 update. *Nucleic Acids Res.* **37**, D717–D722 (2009).
33. Vandin, F., Upfal, E. & Feilcke, S. J. In Proc. 14th Internat. Conf. Res. Comput. Mol. Biol. (ed. Berger, A.) 506–521 (Springer, 2010).
34. Choi, J. H. et al. Zappal, J. and Nodal1 justome loop repeats regulate ovarian tumor growth and adhesion. *Cancer Res.* **68**, 5716–5723 (2008).
35. Farner, H. et al. Targeting the DNA-repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 911–912 (2005).
36. Fong, P. C. et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA1 mutation carriers. *N. Engl. J. Med.* **361**, 123–134 (2009).
37. Veckl, J. et al. BRCA1 C-terminal hypermethylation predicts sensitivity to poly(adenosine diphosphate)-ribose polymerase inhibitors. *J. Clin. Oncol.* **28**, e563–e564 (2010).
38. Yao, T. A. et al. Reply to J. Veckl et al. *J. Clin. Oncol.* **30**, e565–e566 (2010).
39. Mendes-Pereira, A. M. et al. Synthetic lethality targeting of PTEN mutant cells with PARP inhibitors. *EMBO Mol. Med.* **3**, 115–122 (2009).
40. Nakayama, N. et al. Gene amplification CCNE1 is related to poor survival and potential therapeutic target in ovarian cancer. *Cancer* **116**, 2651–2654 (2010).

614 | NATURE | VOL 474 | 30 JUNE 2011

©2011 Macmillan Publishers Limited. All rights reserved.

ARTICLE RESEARCH

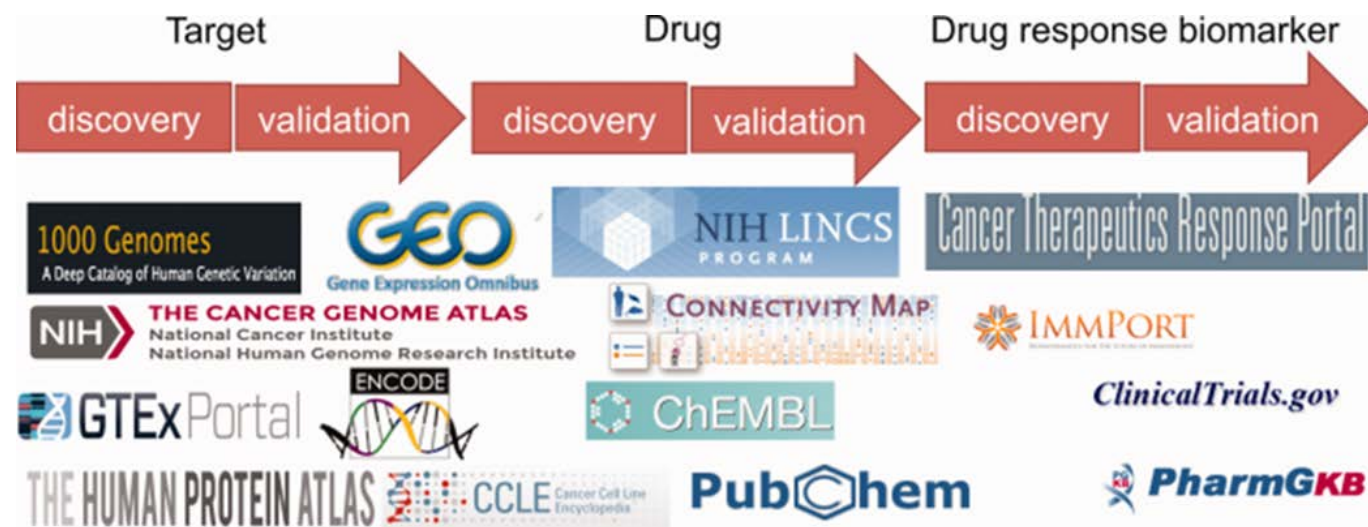
Cancer genome characterization: Broad Institute Dana-Farber Cancer Institute M. Meyerson^{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,287,288,289,290,291,292,293,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,388,389,390,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,443,444,445,446,447,448,449,450,451,452,453,454,455,456,457,458,459,460,461,462,463,464,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479,480,481,482,483,484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,571,572,573,574,575,576,577,578,579,580,581,582,583,584,585,586,587,588,589,590,591,592,593,594,595,596,597,598,599,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,974,975,976,977,978,979,980,981,982,983,984,985,986,987,988,989,990,991,992,993,994,995,996,997,998,999,1000,1001,1002,1003,1004,1005,1006,1007,1008,1009,1010,1011,1012,1013,1014,1015,1016,1017,1018,1019,1020,1021,1022,1023,1024,1025,1026,1027,1028,1029,1030,1031,1032,1033,1034,1035,1036,1037,1038,1039,1040,1041,1042,1043,1044,1045,1046,1047,1048,1049,1050,1051,1052,1053,1054,1055,1056,1057,1058,1059,1060,1061,1062,1063,1064,1065,1066,1067,1068,1069,1070,1071,1072,1073,1074,1075,1076,1077,1078,1079,1080,1081,1082,1083,1084,1085,1086,1087,1088,1089,1090,1091,1092,1093,1094,1095,1096,1097,1098,1099,1100,1101,1102,1103,1104,1105,1106,1107,1108,1109,1110,1111,1112,1113,1114,1115,1116,1117,1118,1119,1120,1121,1122,1123,1124,1125,1126,1127,1128,1129,1130,1131,1132,1133,1134,1135,1136,1137,1138,1139,1140,1141,1142,1143,1144,1145,1146,1147,1148,1149,1150,1151,1152,1153,1154,1155,1156,1157,1158,1159,1160,1161,1162,1163,1164,1165,1166,1167,1168,1169,1170,1171,1172,1173,1174,1175,1176,1177,1178,1179,1180,1181,1182,1183,1184,1185,1186,1187,1188,1189,1190,1191,1192,1193,1194,1195,1196,1197,1198,1199,1200,1201,1202,1203,1204,1205,1206,1207,1208,1209,1210,1211,1212,1213,1214,1215,1216,1217,1218,1219,1220,1221,1222,1223,1224,1225,1226,1227,1228,1229,1230,1231,1232,1233,1234,1235,1236,1237,1238,1239,1240,1241,1242,1243,1244,1245,1246,1247,1248,1249,1250,1251,1252,1253,1254,1255,1256,1257,1258,1259,1260,1261,1262,1263,1264,1265,1266,1267,1268,1269,1270,1271,1272,1273,1274,1275,1276,1277,1278,1279,1280,1281,1282,1283,1284,1285,1286,1287,1288,1289,1290,1291,1292,1293,1294,1295,1296,1297,1298,1299,1300,1301,1302,1303,1304,1305,1306,1307,1308,1309,1310,1311,1312,1313,1314,1315,1316,1317,1318,1319,1320,1321,1322,1323,1324,1325,1326,1327,1328,1329,1330,1331,1332,1333,1334,1335,1336,1337,1338,1339,1340,1341,1342,1343,1344,1345,1346,1347,1348,1349,1350,1351,1352,1353,1354,1355,1356,1357,1358,1359,1360,1361,1362,1363,1364,1365,1366,1367,1368,1369,1370,1371,1372,1373,1374,1375,1376,1377,1378,1379,1380,1381,1382,1383,1384,1385,1386,1387,1388,1389,1390,1391,1392,1393,1394,1395,1396,1397,1398,1399,1400,1401,1402,1403,1404,1405,1406,1407,1408,1409,1410,1411,1412,1413,1414,1415,1416,1417,1418,1419,1420,1421,1422,1423,1424,1425,1426,1427,1428,1429,1430,1431,1432,1433,1434,1435,1436,1437,1438,1439,1440,1441,1442,1443,1444,1445,1446,1447,1448,1449,1450,1451,1452,1453,1454,1455,1456,1457,1458,1459,1460,1461,1462,1463,1464,1465,1466,1467,1468,1469,1470,1471,1472,1473,1474,1475,1476,1477,1478,1479,1480,1481,1482,1483,1484,1485,1486,1487,1488,1489,1490,1491,1492,1493,1494,1495,1496,1497,1498,1499,1500,1501,1502,1503,1504,1505,1506,1507,1508,1509,1510,1511,1512,1513,1514,1515,1516,1517,1518,1519,1520,1521,1522,1523,1524,1525,1526,1527,1528,1529,1530,1531,1532,1533,1534,1535,1536,1537,1538,1539,1540,1541,1542,1543,1544,1545,1546,1547,1548,1549,1550,1551,1552,1553,1554,1555,1556,1557,1558,1559,1560,1561,1562,1563,1564,1565,1566,1567,1568,1569,1570,1571,1572,1573,1574,1575,1576,1577,1578,1579,1580,1581,1582,1583,1584,1585,1586,1587,1588,1589,1590,1591,1592,1593,1594,1595,1596,1597,1598,1599,1600,1601,1602,1603,1604,1605,1606,1607,1608,1609,1610,1611,1612,1613,1614,1615,1616,1617,1618,1619,1620,1621,1622,1623,1624,1625,1626,1627,1628,1629,1630,1631,1632,1633,1634,1635,1636,1637,1638,1639,1640,1641,1642,1643,1644,1645,1646,1647,1648,1649,1650,1651,1652,1653,1654,1655,1656,1657,1658,1659,1660,1661,1662,1663,1664,1665,1666,1667,1668,1669,1670,1671,1672,1673,1674,1675,1676,1677,1678,1679,1680,1681,1682,1683,1684,1685,1686,1687,1688,1689,1690,1691,1692,1693,1694,1695,1696,1697,1698,1699,1700,1701,1702,1703,1704,1705,1706,1707,1708,1709,1710,1711,1712,1713,1714,1715,1716,1717,1718,1719,1720,1721,1722,1723,1724,1725,1726,1727,1728,1729,1730,1731,1732,1733,1734,1735,1736,1737,1738,1739,1740,1741,1742,1743,1744,1745,1746,1747,1748,1749,1750,1751,1752,1753,1754,1755,1756,1757,1758,1759,1760,1761,1762,1763,1764,1765,1766,1767,1768,1769,1770,1771,1772,1773,1774,1775,1776,1777,1778,1779,1780,1781,1782,1783,1784,1785,1786,1787,1788,1789,1790,1791,1792,1793,1794,1795,1796,1797,1798,1799,1800,1801,1802,1803,1804,1805,1806,1807,1808,1809,1810,1811,1812,1813,1814,1815,1816,1817,1818,1819,1820,1821,1822,1823,1824,1825,1826,1827,1828,1829,1830,1831,1832,1833,1834,1835,1836,1837,1838,1839,1840,1841,1842,1843,1844,1845,1846,1847,1848,1849,1850,1851,1852,1853,1854,1855,1856,1857,1858,1859,1860,1861,1862,1863,1864,1865,1866,1867,1868,1869,1870,1871,1872,1873,1874,1875,1876,1877,1878,1879,1880,1881,1882,1883,1884,1885,1886,1887,1888,1889,1890,1891,1892,1893,1894,1895,1896,1897,1898,1899,1900,1901,1902,1903,1904,1905,1906,1907,1908,1909,1910,1911,1912,1913,1914,1915,1916,1917,1918,1919,1920,1921,1922,1923,1924,1925,1926,1927,1928,1929,1930,1931,1932,1933,1934,1935,1936,1937,1938,1939,1940,1941,1942,1943,1944,1945,1946,1947,1948,1949,1950,1951,1952,1953,1954,1955,1956,1957,1958,1959,1960,1961,1962,1963,1964,1965,1966,1967,1968,1969,1970,1971,1972,1973,1974,1975,1976,1977,1978,1979,1980,1981,1982,1983,1984,1985,1986,1987,1988,1989,1990,1991,1992,1993,1994,1995,1996,1997,1998,1999,2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,2021,2022,2023,2024,2025,2026,2027,2028,2029,2030,2031,2032,2033,2034,2035,2036,2037,2038,2039,2040,2041,2042,2043,2044,2045,2046,2047,2048,2049,2050,2051,2052,2053,2054,2055,2056,2057,2058,2059,2060,2061,2062,2063,2064,2065,2066,2067,2068,2069,2070,2071,2072,2073,2074,2075,2076,2077,2078,2079,2080,2081,2082,2083,2084,2085,2086,2087,2088,2089,2090,2091,2092,2093,2094,2095,2096,2097,2098,2099,2100,2101,2102,2103,2104,2105,2106,2107,2108,2109,2110,2111,2112,2113,2114,2115,2116,2117,2118,2119,2120,2121,2122,2123,2124,2125,2126,2127,2128,2129,2130,2131,2132,2133,2134,2135,2136,2137,2138,2139,2140,2141,2142,2143,2144,2145,2146,2147,2148,2149,2150,2151,2152,2153,2154,2155,2156,2157,2158,2159,2160,2161,2162,2163,2164,2165,2166,2167,2168,2169,2170,2171,2172,2173,2174,2175,2176,2177,2178,2179,2180,2181,2182,2183,2184,2185,2186,2187,2188,2189,2190,2191,2192,2193,2194,2195,2196,2197,2198,2199,2200,2201,2202,2203,2204,2205,2206,2207,2208,2209,2210,2211,2212,2213,2214,2215,2216,2217,2218,2219,2220,2221,2222,2223,2224,2225,2226,2227,2228,2229,2230,2231,2232,2233,2234,2235,2236,2237,2238,2239,2240,2241,2242,2243,2244,2245,2246,2247,2248,2249,2250,2251,2252,2253,2254,2255,2256,2257,2258,2259,2260,2261,2262,2263,2264,2265,2266,2267,2268,2269,2270,2271,2272,2273,2274,2275,2276,2277,2278,2279,2280,2281,2282,2283,2284,2285,2286,2287,2288,2289,2290,2291,2292,2293,2294,2295,2296,2297,2298,2299,2300,2301,2302,2303,2304,2305,2306,2307,2308,2309,2310,2311,2312,2313,2314,2315,2316,2317,2318,2319,2320,2321,2322,2323,2324,2325,2326,2327,2328,2329,2330,2331,2332,2333,2334,2335,2336,2337,2338,2339,2340,2341,2342,2343,2344,2345,2346,2347,2348,2349,2350,2351,2352,2353,2354,2355,2356,2357,2358,2359,2360,2361,2362,2363,2364,2365,2366,2367,2368,2369,2370,2371,2372,2373,23}

Data sharing policy

- The concepts of data sharing and open data are becoming increasingly important in science
- Funding bodies, journals and societies are now encouraging or mandating data sharing (usually the raw data)
- Sharing data publicly is an important way of improving reproducibility and showing that researchers are confident in their work
- Studies with raw data shared in a repository also receive more citations than those without publicly available data

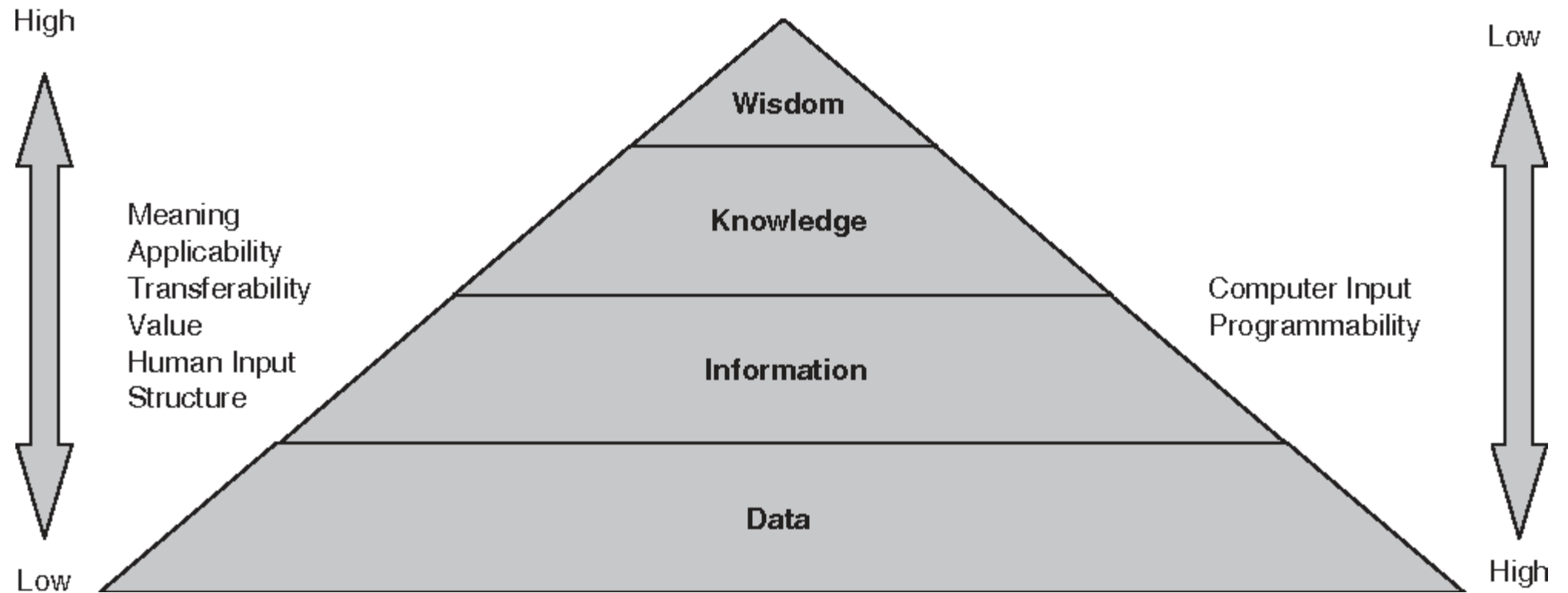
- But raw -omics data are hard to analyse, so many platforms gather the publicly available data, thoroughly analyze it, curate it and share it in a user friendly format

Leveraging Public Databases to Identify Actionable Targets



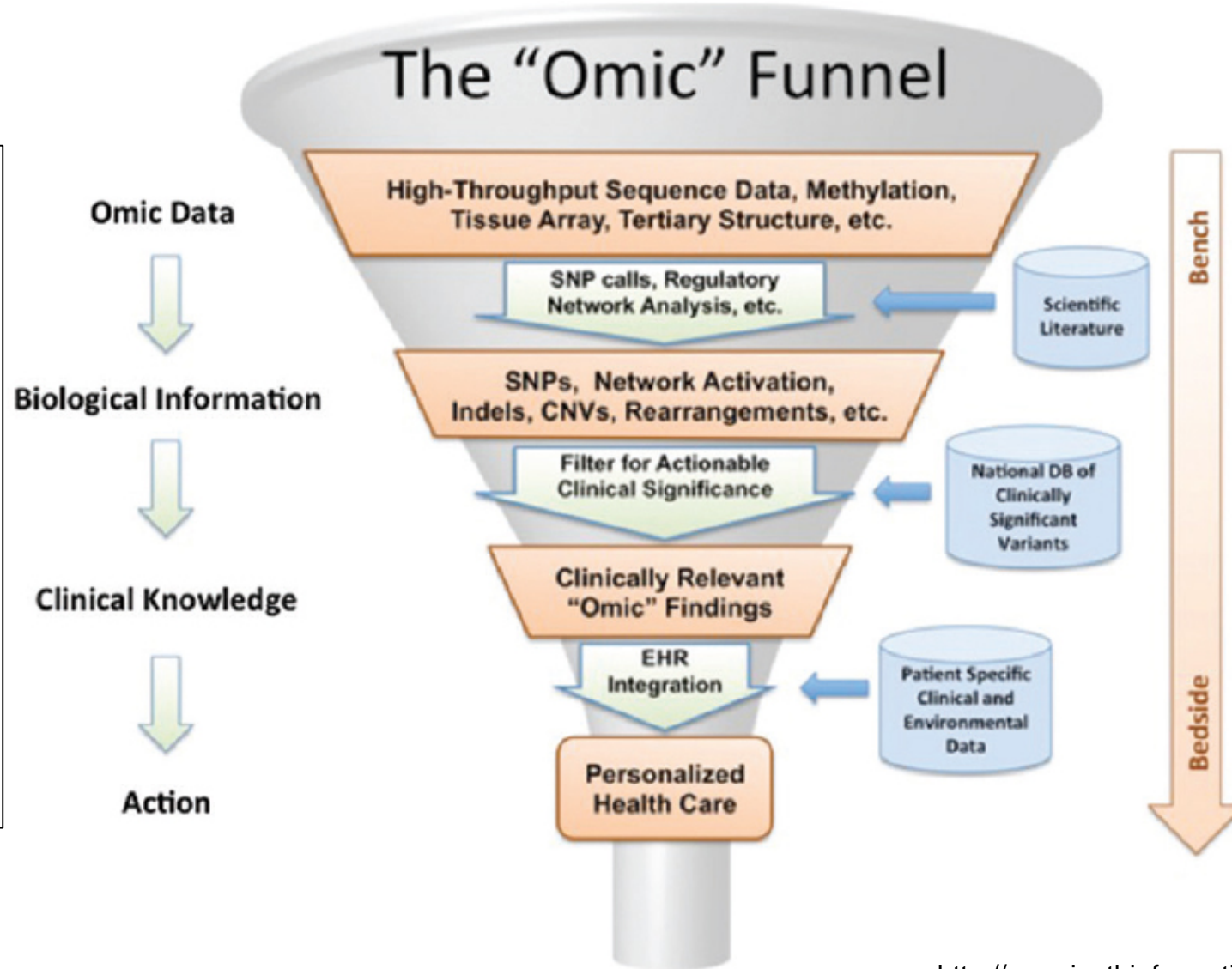
DIKW pyramide

„Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom.“ – Clifford Stoll



What is the aim of OMICS technologies

In Genomics:
 Sequence of 3 billions letters in .txt file
 ↓
 information where the individual's genome varies from reference sequence
 ↓
CYP2C9 or *TPMT* genotype, which has known pharmacogenomic associations
 ↓
 individualize the dose of a new warfarin prescription



The DIKW pyramid metaphor:

"know-nothing" (Data)
 ↓
 "know-what" (Information)
 ↓
 "know-how" (Knowledge)
 ↓
 "Know-why" (Wisdom)

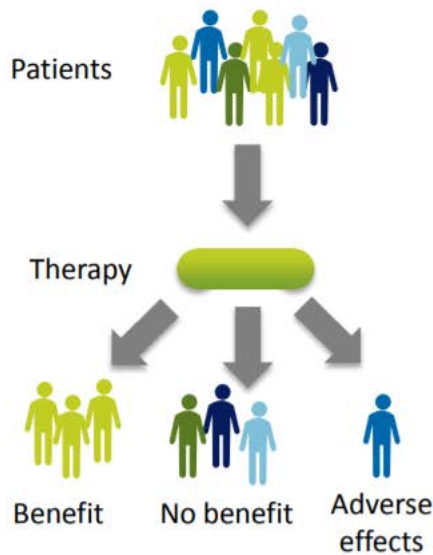
Zeleny (2005)

What is personalized health care?

Personalized medicine, sometimes referred to as *precision* or *individualized* medicine, is an emerging field of medicine that uses diagnostic tools to identify specific biological markers, often genetic, to help assess which medical treatments and procedures will be best for each patient.

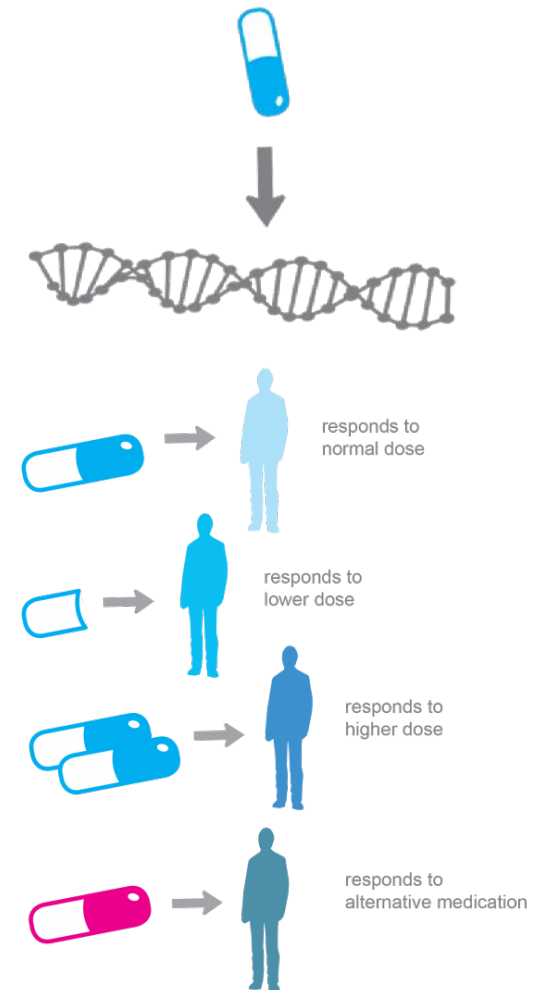
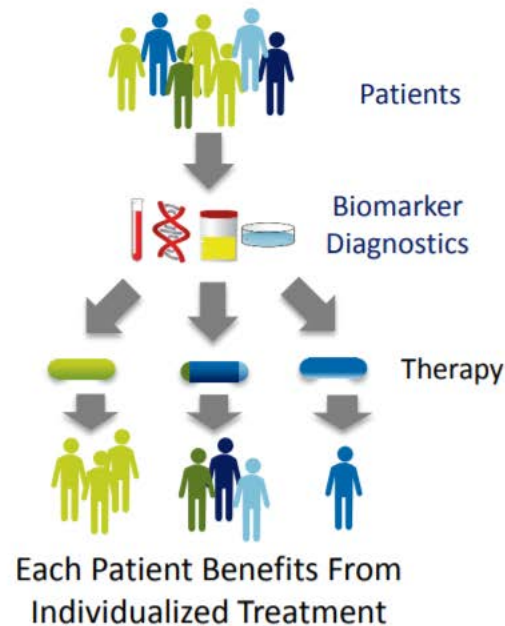
Without Personalized Medicine:

Some Benefit, Some Do Not



With Personalized Medicine:

Each Patient Receives the Right Medicine For Them



Value of personalized medicine



\$5 billion

Estimated annual cost of wasted prescription drugs in the US.

\$3 billion

Estimated cost of wasted hospital cancer drugs.



27%

of all NMEs approved by the FDA in 2016 are personalized medicines.



50%

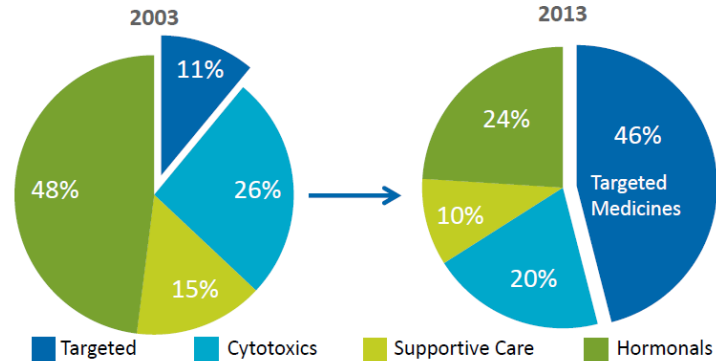
of the personalized medicines approved by the FDA in 2016 are oncology drugs.

<https://invivo.pharmaintelligence.informa.com/>

Oncology is on the Leading Edge of Personalized Medicine

In ten years, cancer patients have seen a four-fold increase in their personalized medicine treatment options.

*Breakdown of Oncology Treatment Modalities, Global Market share 2003-2013**



Personalized Medicine Can Create Efficiencies in the Health Care System

Breast Cancer



Reduction in chemotherapy use would occur

If women with breast cancer receive a genetic test of their tumor prior to treatment

Metastatic Colorectal Cancer



In annual health care cost savings would be realized

If patients with metastatic colorectal cancer receive a genetic test for the KRAS gene prior to treatment

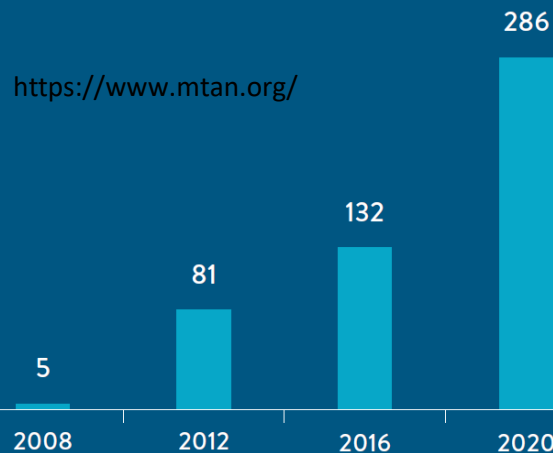
Stroke



Strokes could be prevented each year

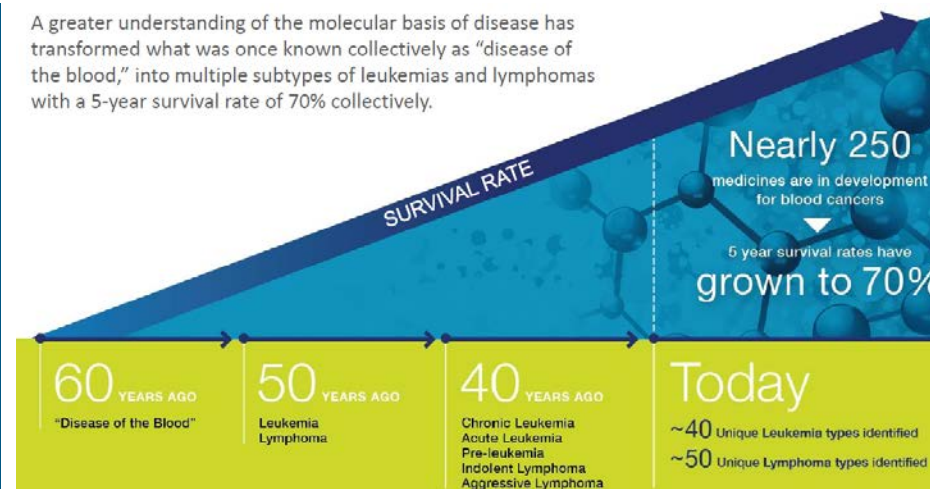
If a genetic test is used to properly dose blood thinners

PERSONALIZED MEDICINES ON THE MARKET

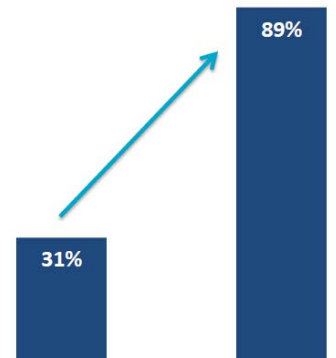


<https://www.mtan.org/>

A greater understanding of the molecular basis of disease has transformed what was once known collectively as "disease of the blood," into multiple subtypes of leukemias and lymphomas with a 5-year survival rate of 70% collectively.



5-Year Survival Rates for CML Patients Nearly Triple After Introduction of Imatinib

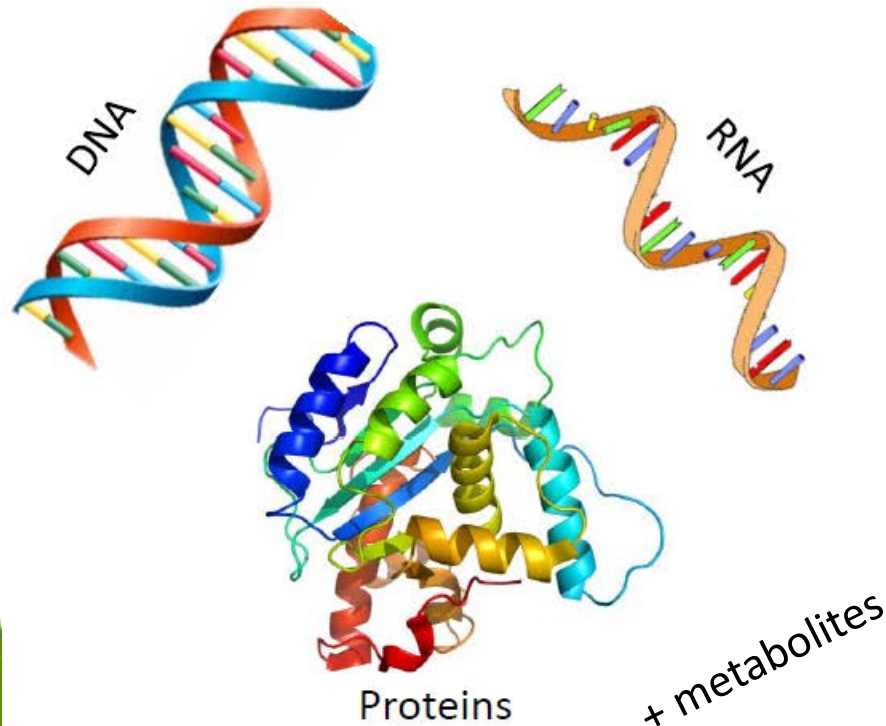


And many more examples, see <http://www.personalizedmedicinecoalition.org> for more detailed information on PM

What is a biomarker?

A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, disease processes, or biological responses to a therapeutic intervention. Biomarkers can be used to reduce uncertainty and guide clinical care.

Molecular Biomarkers Can Include:



Biomarkers Help Inform Medical Decisions:

- Prevention measures?
- Which diagnosis?
- Treat or don't treat?
- What dose?

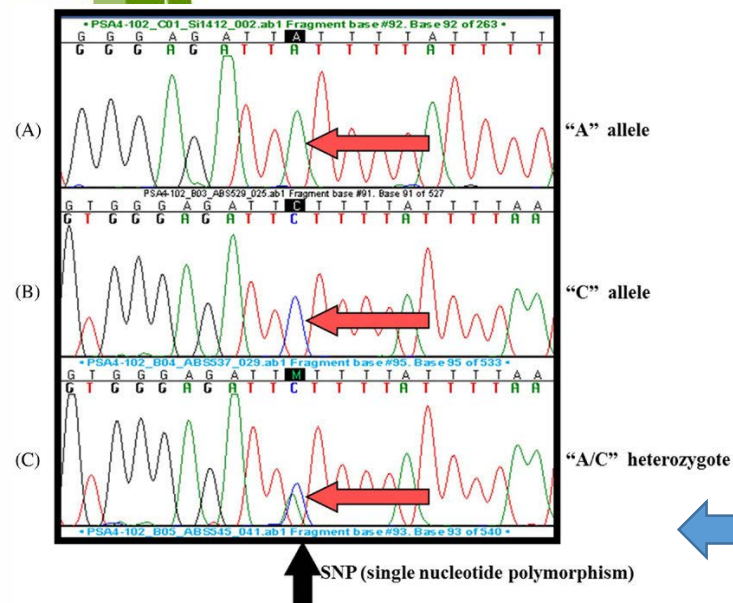
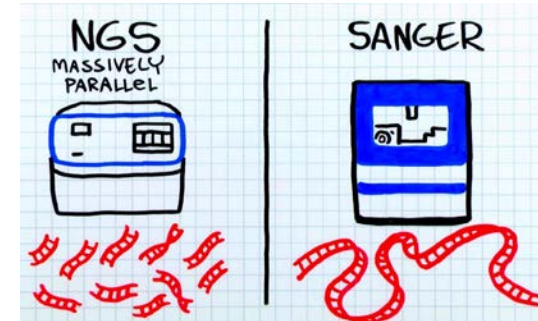
How Do You Detect a Biomarker?

- Diagnostics
 - Blood draw
 - Microscopic analysis
 - Gene sequencing
 - Biopsy
 - Protein analysis

-OMICS technologies and their integration is crucial for biomarker discovery and validation

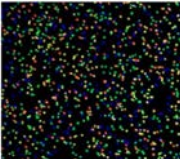
Sanger vs next generation sequencing

- Sanger sequencing
 - <https://www.youtube.com/watch?v=e2G5zx-OJlw>
- Next generation sequencing (Illumina is shown as an example)
 - <https://www.youtube.com/watch?v=9YxExTSwgPM>



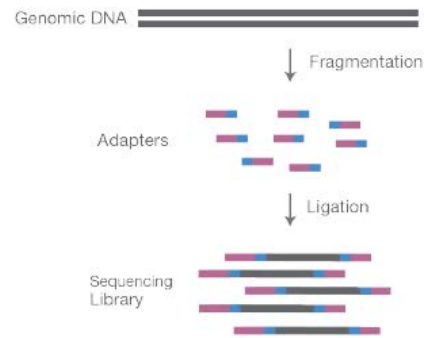
Sanger Sequencing

Illumina Sequencing

Sanger Sequencing		Illumina Sequencing	
Advantages	Disadvantages	Advantages	Disadvantages
Lowest error rate (1.5%)	High cost per base	Low error rate	Must run at very large scale
Long read length (~750 bp)	Long time to generate data	Lowest cost per base	Short read length (50-75 bp)
Can target a primer	Need for cloning	Tons of data	Runs take multiple days
Used to confirm NGS results	Amount of data per run	 An image of hundreds of extended molecules	High startup costs
Seeing is believing			De Novo assembly difficult

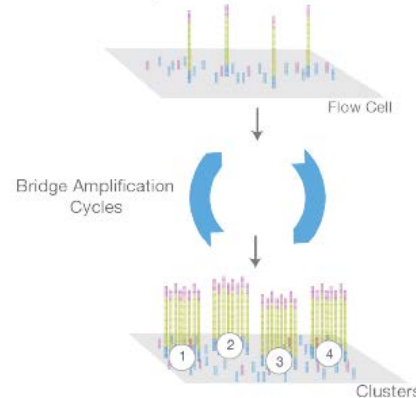
Illumina NGS overview

A. Library Preparation



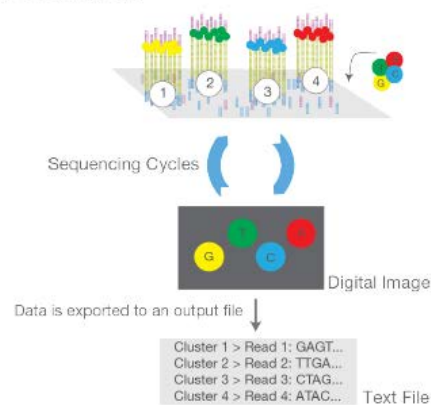
NGS library is prepared by fragmenting a gDNA sample and ligating specialized adapters to both fragment ends.

B. Cluster Amplification



Library is loaded into a flow cell and the fragments are hybridized to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.

C. Sequencing



Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated "n" times to create a read length of "n" bases.

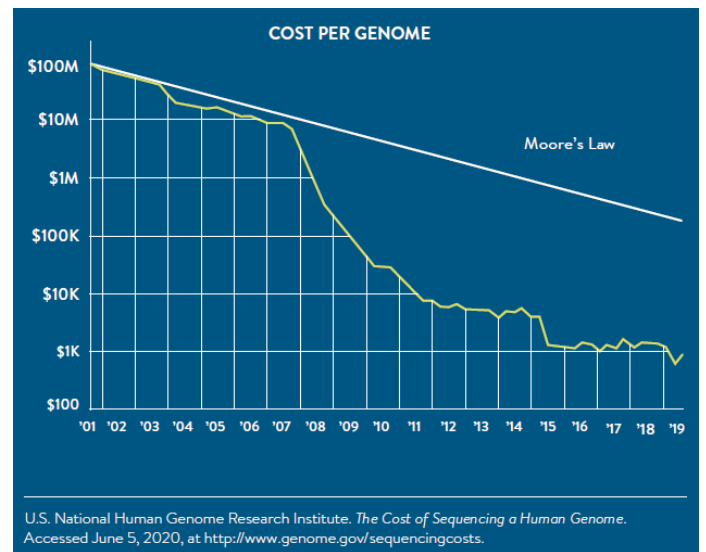
D. Alignment and Data Analysis



Reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified.



- www.illumina.com/technology/next-generation-sequencing.html



DTC (direct-to-customer) genetic testing

ancestry

FREE TRIAL [SIGN IN >](#)



Give the gift that has connected
20 million members to a deeper family story.

ONLY **\$59*** ~~\$99~~

[Give AncestryDNA®](#)

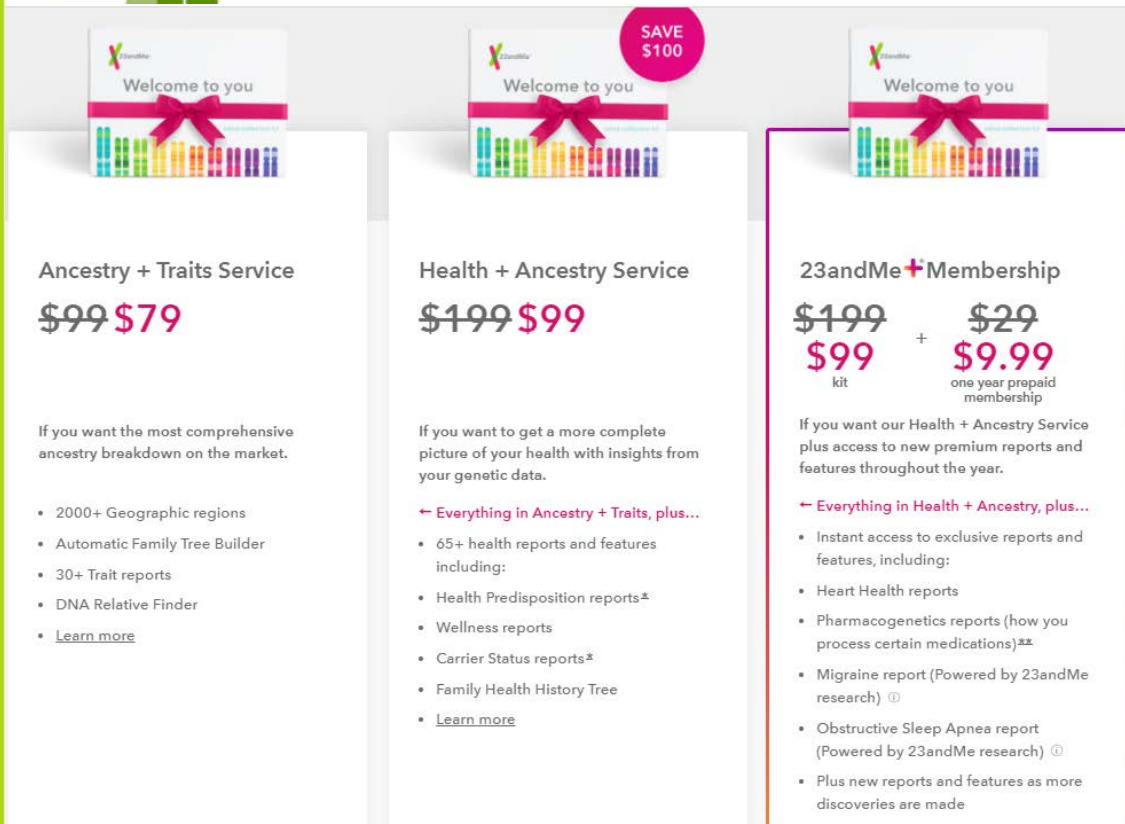
*Offer ends 11/21. Excludes taxes and shipping.

 Build a family tree to see your story emerge.

[Learn more](#)

Genotyping vs Sequencing

- **Genotyping** - determining which genetic variants an individual possesses through a variety of different methods, especially genotyping chips (based mostly on SNPs – single nucleotide polymorphisms)
 - cheap, but require prior identification of the variants of interest



The screenshot displays three service tiers from 23andMe, each with a 'Welcome to you' card featuring a DNA microarray image. The first tier, 'Ancestry + Traits Service', is priced at \$99 (reduced from \$99) and includes 2000+ geographic regions, an automatic family tree builder, 30+ trait reports, and a DNA relative finder. The second tier, 'Health + Ancestry Service', is priced at \$199 (reduced from \$199) and includes 65+ health reports, predisposition reports, wellness reports, carrier status reports, and a family health history tree. The third tier, '23andMe+ Membership', is priced at \$199 (reduced from \$199) plus a \$29 kit and a \$9.99 one-year prepaid membership, offering instant access to exclusive reports and features like heart health, pharmacogenetics, migraine, and sleep apnea reports.

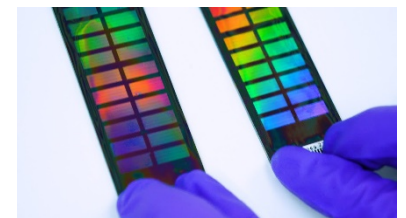
Service	Price	Key Features
Ancestry + Traits Service	\$99 (was \$99)	2000+ Geographic regions, Automatic Family Tree Builder, 30+ Trait reports, DNA Relative Finder
Health + Ancestry Service	\$199 (was \$199)	65+ health reports and features including: Health Predisposition reports, Wellness reports, Carrier Status reports, Family Health History Tree
23andMe+ Membership	\$199 (was \$199) + \$29 kit + \$9.99 one year prepaid membership	Instant access to exclusive reports and features, including: Heart Health reports, Pharmacogenetics reports, Migraine report, Obstructive Sleep Apnea report, Plus new reports and features as more discoveries are made

Methods

We use genotyping technology to look at specific genetic variants in the genome that can be most informative about an individual's health and ancestry.

Unlike sequencing which analyses all nucleotides in a gene to identify changes, genotyping detects specific known variants within the genome. 23andMe uses a custom Illumina HumanOmniExpress-24 format chip that analyses approximately half a million variants. This custom chip has been designed to include variants:

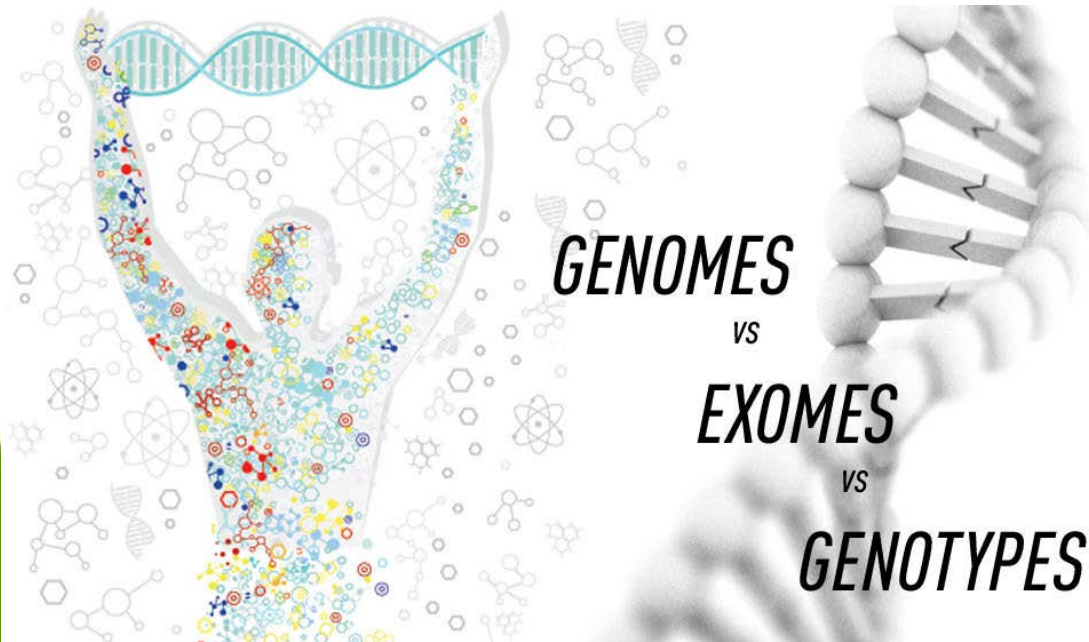
- In medically relevant genes
- Involved in drug metabolism, efficacy and side effects
- With known disease associations
- Associated with traits
- Used to assign genetic ancestry and ethnicity



<https://www.23andme.com/>

How SNP genotyping works

- https://www.youtube.com/watch?v=Naona1y_I2U
- For more information see YouTube Channel Useful Genetics:
<https://www.youtube.com/channel/UCtXCrx28msMBQ-vFUIOIReA>

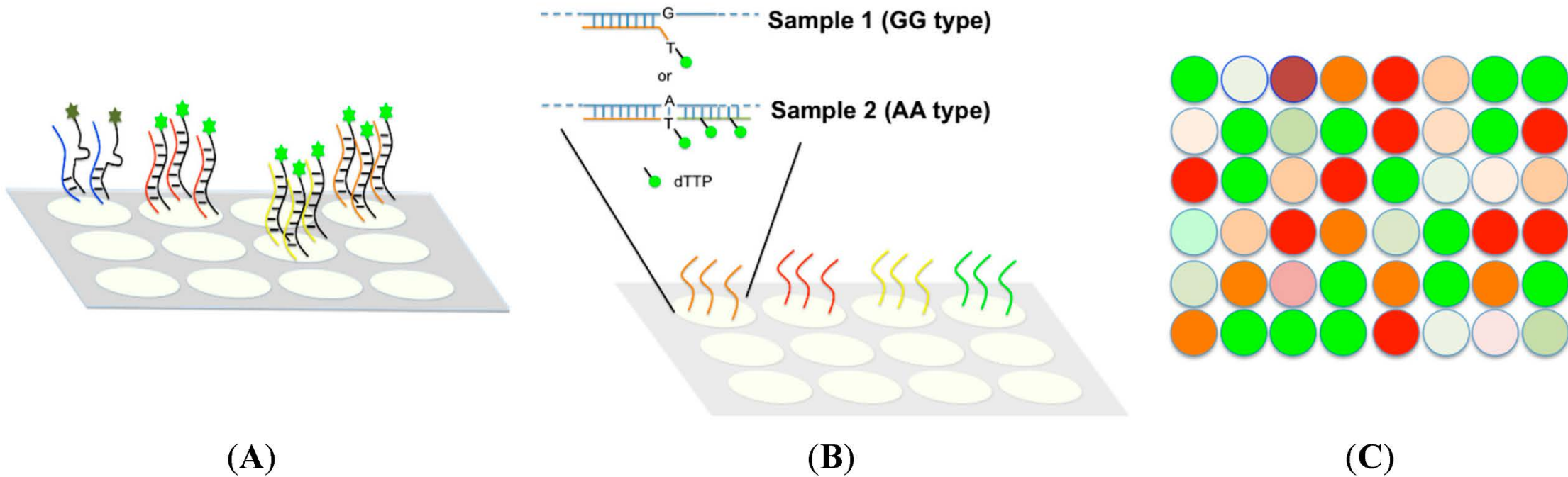


<https://www.jax.org/news-and-insights/jax-blog/2016/september/genomes-versus-exomes-versus-genotypes>

SNP - Single nucleotide polymorphisms

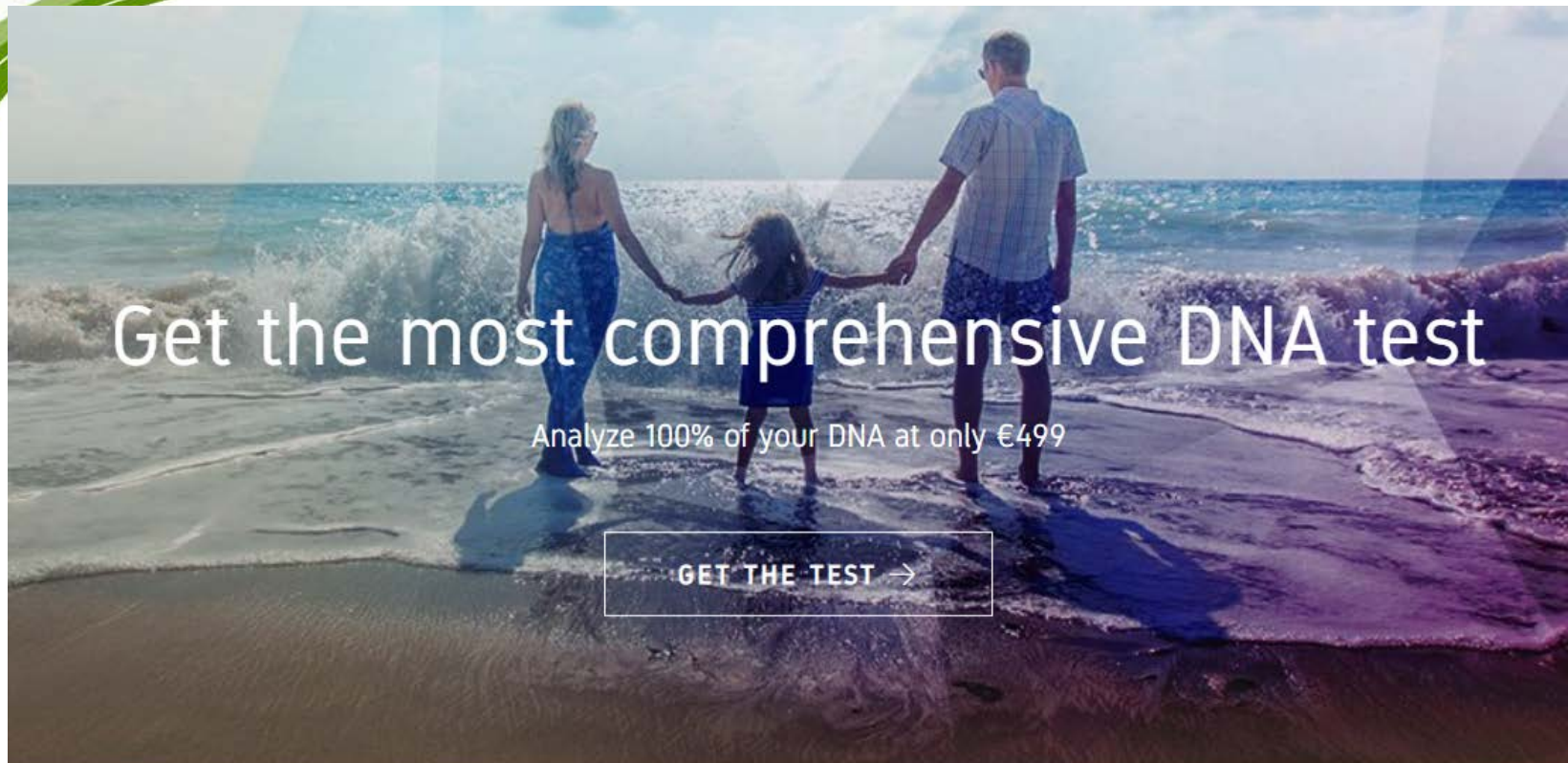
- the most common type of genetic variation
- occur almost once in every 1,000 nucleotides on average, 4 to 5 million SNPs in a person's genome
- may be unique or occur in many individuals; scientists have found more than 100 million SNPs in populations around the world
- most commonly in non-coding DNA
- can act as biological markers, helping locate genes associated with disease
- most SNPs have no effect on health or development
- some SNPs have proven to be very important in the study of human health.
- may help predict an individual's response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases.
- SNPs can also be used to track the inheritance of disease genes within families

How SNP genotyping works



There are two types of microarray commonly used in multiplexing SNP analysis: allele-specific oligonucleotide (ASO) hybridization and allele-specific primer (ASP) extension. **(A)** ASO hybridization: The allele-specific oligonucleotide for every SNP is synthesized and separately immobilized onto the glass plate. Fluorescence labeled targets containing SNP sites are produced from a PCR reaction and plotted separately into each well to conduct the hybridization reaction. The mismatched base pair between target and oligonucleotide can decrease the binding strength with the fluorescence-labeled target removed after a stringent washing. A fluorescence signal is detected on a perfectly matched base pair; **(B)** Allele-specific primer (ASP) extension: The specific primer for SNP location is designed and separately immobilized onto a microarray. A different fluorescence labeled dNTP is individually used in an extension reaction. The extended fragment showing fluorescence signal can only be found when the 3' end of primer pair is perfectly matched (AA type in this case) in contrast to the mismatched primer pair (GG type in this case); **(C)** The SNP genotype can be determined according to fluorescent intensity from the products/target DNA. <https://doi.org/10.3390/microarrays4040570>

DTC genome sequencing as popular demand



Get the most comprehensive DNA test

Analyze 100% of your DNA at only €499

[GET THE TEST →](#)

Coverage (or depth) in sequencing

```
Read 1: CGGATTACGTGGACCATG (read length of 18)
Read 2: ATTACGTGGACCATGAATTGCTGACA
Read 3: ACCATGAATTGCTGACATTCGTCA
Read 4: TGAATTGCTGACATTCGTCA
Depth: 1112222222233334433333333332222221
```

WHAT YOU GET

Dante Labs analyzes 100% of your DNA, so that we can give you reports on predispositions on any genetic disease. You will receive easy reports for you and your doctor, as well as raw data to explore.



SALE

My Full DNA: Whole Genome Sequencing with mtDNA

€449.00 EUR ~~€850.00-EUR~~
YOU SAVE €401.00 EUR

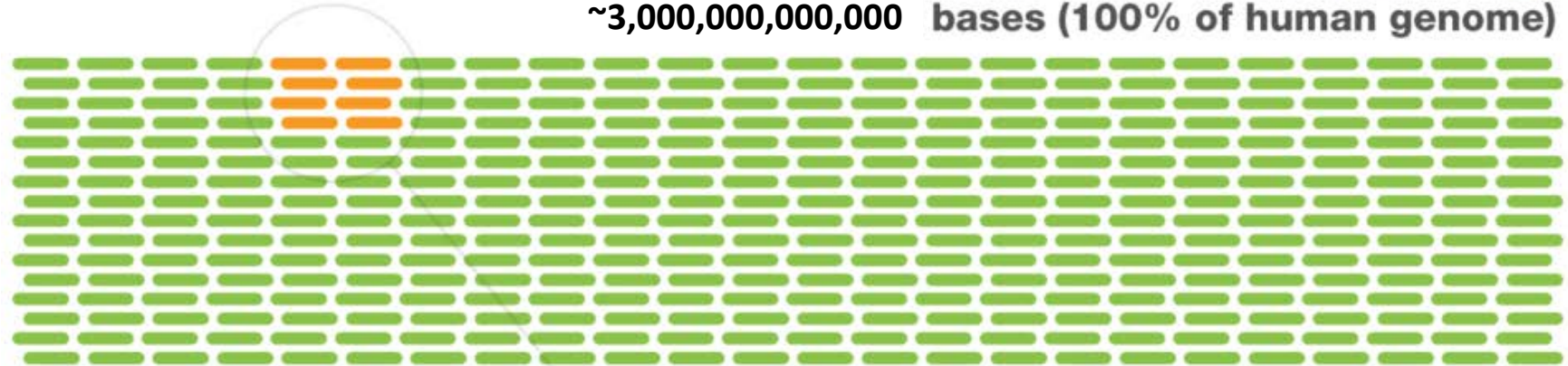
www.dantelabs.com

Sequencing – WGS and WES

- Determining the exact DNA sequence

Whole Genome Sequencing

~3,000,000,000,000 bases (100% of human genome)

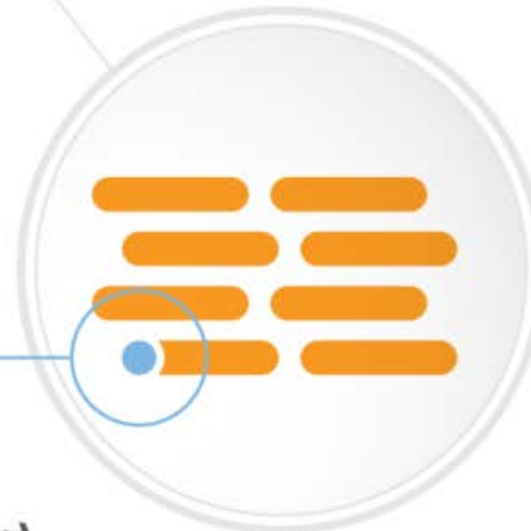


Whole Exome Sequencing

~60,000,000 bases
(~2% of human genome)

Large Scale Genotyping

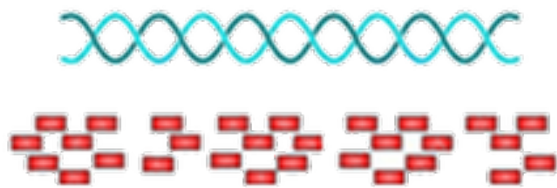
~1,000,000 bases
(~0.03% of human genome)



“Non-coding DNA” was long thought of as junk DNA, but as we understand more about our genetics we now know these regions play a hugely important role in regulating the coding portions of our DNA. Our understanding of these regions and their interactions is relatively poor compared to our knowledge of the DNA coding regions.

Genomes vs exomes vs genotypes

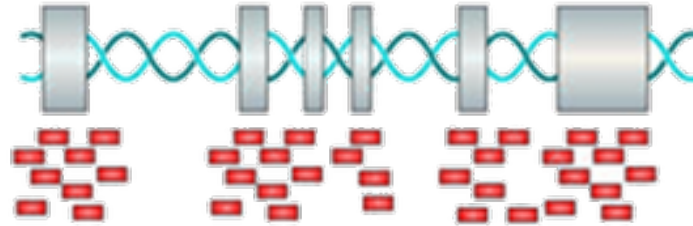
WGS Whole genome sequencing



- Sequencing region : whole genome
- Sequencing Depth: >30X
- Covers everything – can identify all kinds of variants including SNPs, INDELs and SV.

- Results are sometimes challenging to interpret

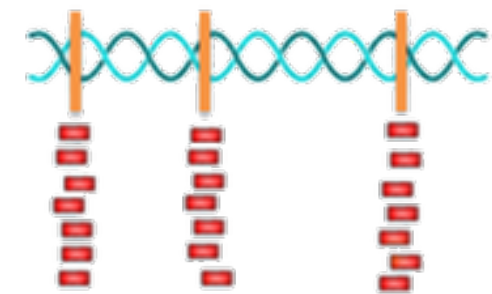
WES Whole exome sequencing



- Sequencing region: whole exome
- Sequencing Depth : >50X ~ 100X
- Identify all kinds of variants including SNPs, INDELs and SV in coding region.
- Cost effective

- Good alternative to WGS in terms of clinical use

Hotspot sequencing = Targeted sequencing



- Sequencing region: specific regions (could be customized)
- Sequencing Depth : >500X
- Identify all kinds of variants including SNPs, INDELs in specific regions
- Most Cost effective

- Most sensitive – able to detect rare tumor cells in a biopsy

What to expect

- Genetic testing provided by most of the companies is moreless for fun (ancestry, health and wellness, nutrigenetics, skincare, sports,...)

How it Works

Fitness Report for Longevity
The Fitness report underlines the genetic predispositions to various aspects related to physics and sports activity, to obtain the best performance in the shortest possible time.

Nutrigenomic Report for Longevity
The Nutrigenomics report shows all the relationships between specific foods and genes, in order to support the creation of the most effective diet while respecting individual predispositions.

LIVELONGER WHOLE GENOME TEST
SALIVA COLLECTION KIT
Dante Health
ACTIONABLE INSIGHTS TO LIVE A LONG HEALTHY LIFE
EXPEDITED SHIPPING LAB IN

<https://dantelabs.com/products/livelonger-genome-test>

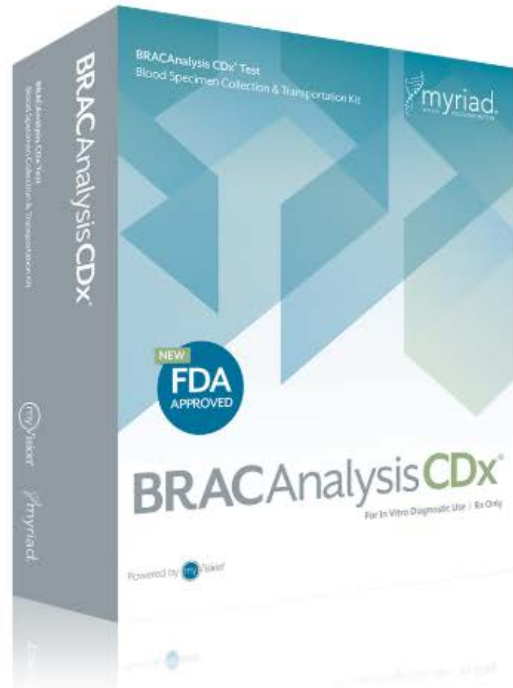
- More expensive, and complete, sequencing like the one provided by Illumina can be used for medical investigation
- Do not expect your genome sequencing to tell you how long is your life expectation, whether you are likely to get cancer and so on
- So far our knowledge on the “implication” of the genome are quite limited
- What we can already do in health care is to look at the genome once you have been diagnosed a specific ailment and look for specific genes that would make one cure more effective than another (this has become normal practice in some form of cancer cure)

Example of genetic testing in clinical practise

- *BRCA* genes testing for PARP inhibitor treatment

BRACAnalysis CDx[®] Ovarian Cancer

Overview



Mutations in *BRCA1* or *BRCA2* cause Hereditary Breast and Ovarian Syndrome (HBOC). Now mutations in the *BRCA1* and *BRCA2* genes provide an indication for treatment with Lynparza™ (olaparib) for patients with ovarian cancer. Specifically, BRACAnalysis CDx[®] is the only FDA-approved laboratory developed test approved to be used to inform treatment decisions for the PARP inhibitor, Lynparza. A positive BRACAnalysis CDx result in patients with ovarian cancer is also associated with enhanced progression-free survival (PFS) from Zejula™ (niraparib) maintenance therapy.^{1,2,3}

Learn More

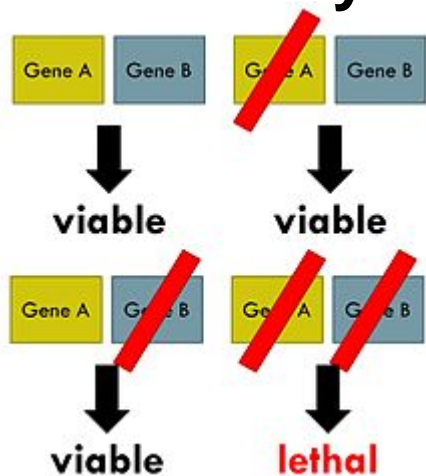
Order BRACAnalysis CDx

- More info: <https://www.youtube.com/watch?v=ilwMGRH276M>

PARP inhibitors

- In December 2014, the drug olaparib (Lynparza) became the first of a new class of treatments known as PARP (poly(ADP-ribose)polymerase) inhibitors to be licensed for clinical use, heralding in a new era for personalised, targeted treatment—and turning the promise of ‘synthetic lethality’ into reality.

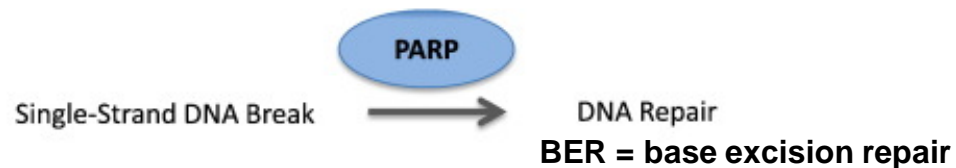
Synthetic lethality concept



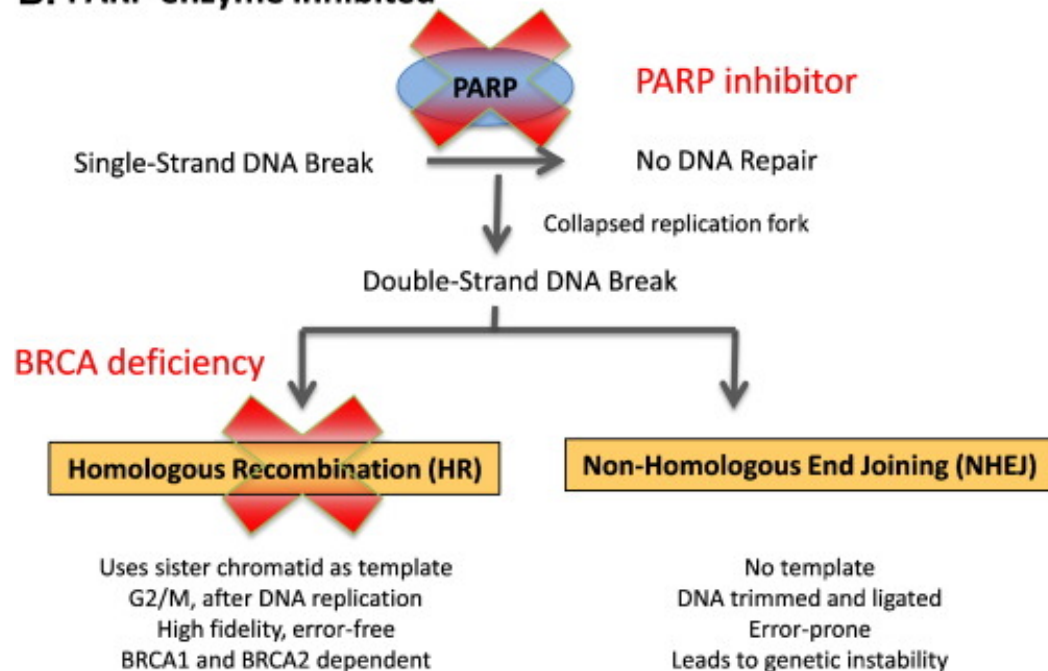
More info on PARPi:

<https://www.youtube.com/watch?v=mgW30YyaJz4>

A. Functioning PARP enzyme



B. PARP enzyme inhibited



C. Deficiency in HR and BER together lead to synthetic lethality

Condition	HR	BER	Outcome
Normal cells	+	+	Viable
BRCA deficient	-	+	Viable
Normal cells, PARP inhibitor	+	-	Viable
BRCA deficient, PARP inhibitor	-	-	Cell Death

<https://doi.org/10.1016/j.ygyno.2015.02.017>

The Present and Future of Genome Sequencing

- Genomics England - **100,000 patients** with rare diseases, their families, and cancer patients
- Precision Medicine Initiative (PMI) **1-million-volunteer** health study, data including genetics and lifestyle factors
- GenomeAsia **100K** - genomic data for Asian populations
- ... a many more initiatives
- How to handle such huge amount of data and the ethical implications?
- In the US, the Genetic Information Nondiscrimination Act (2008) but mostly no act in other countries and somewhat grey legal position in Europe



<https://labiotech.eu/features/genome-sequencing-review-projects/>

Time for
a BREAK



COSMIC: Catalogue of Somatic Mutations in Cancer



<https://cancer.sanger.ac.uk/cosmic/>



Projects ▾ Data ▾ Tools ▾ News ▾ Help ▾ About ▾ Genome Version ▾ Search COSMIC... **SEARCH**

Login ▾

Terms and Conditions have been updated and include important changes. Please check the [Licensing](#) page for details.

COSMIC v94, released 28-MAY-21

COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below.

eg Braf, COLO-829, Carcinoma, V600E, BRCA-UK, Campbell

SEARCH

Projects

COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:



[COSMIC](#)

The core of COSMIC, an expert-curated database of somatic mutations



[Cell Lines Project](#)

Mutation profiles of over 1,000 cell lines used in cancer research



[COSMIC-3D](#)

An interactive view of cancer mutations in the context of 3D structures



[Cancer Gene Census](#)

A catalogue of genes with mutations that are causally implicated in cancer



[Cancer Mutation Census](#)

Classification of genetic variants driving cancer



[Actionability](#)

Mutations actionable in precision oncology

Data curation

- How to use COSMIC database:

<https://www.youtube.com/watch?v=2FD5RabgK6o>, <https://www.youtube.com/watch?v=k477uAiKx74>

COSMIC News

[Follow @cosmic_sanger](#)



Digging for rare finds - three breast cancer publications to keep a watch for in V95

COSMIC V95 will have a focus on rare female cancers, including rare breast cancers. Our latest blog takes a closer look at three of these. [More...](#)



Curating the future of precision oncology: An interview with Steve Jupe

Lean about the curation process, background to Actionability, and innovative uses of COSMIC data in our interview with Steve Jupe. [More...](#)



COSMIC Release v94 is live!

a focus on rare lung cancers and rare pancreatic cancers, and curation of somatic mutations in 12 hallmark apoptosis genes. Along with this, 9 cancer hallmark genes data are also updated. Find out more before exploring the v94 release. [More...](#)

Tools

- [Cancer Browser](#) — browse COSMIC data by tissue type and histology
- [Genome Browser](#) — browse the human genome with COSMIC annotations
- [GA4GH Beacon](#) — access COSMIC data through the [GA4GH Beacon Project](#)
- [COSMIC in BigQuery](#) — search COSMIC via the [ISB Cancer Genomics Cloud](#)

TCGA: The Cancer Genome Atlas



1-800-4-CANCER Live Chat Publications Dictionary

ABOUT CANCER CANCER TYPES RESEARCH GRANTS & TRAINING NEWS & EVENTS ABOUT NCI search

Home > About NCI > NCI Organization > CCG > Research > Structural Genomics



TCGA

Program History +

TCGA Cancers Selected for Study

Publications by TCGA

Using TCGA +

Contact

The Cancer Genome Atlas Program

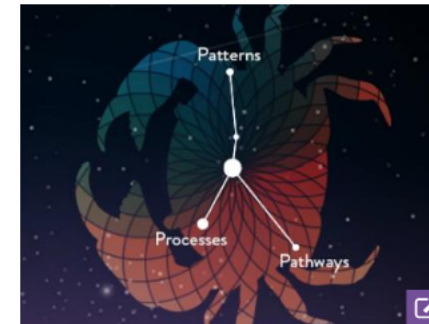
The Cancer Genome Atlas (TCGA), a landmark [cancer genomics](#) program, molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types. This joint effort between the National Cancer Institute and the National Human Genome Research Institute began in 2006, bringing together researchers from diverse disciplines and multiple institutions.

Over the next dozen years, TCGA generated over 2.5 petabytes of genomic, epigenomic, transcriptomic, and proteomic data. The data, which has already led to improvements in our ability to diagnose, treat, and prevent cancer, will remain [publicly available](#) for anyone in the research community to use.



TCGA Outcomes & Impact

TCGA has changed our understanding of cancer, how research is conducted, how the disease is treated in the clinic, and more.



TCGA's PanCancer Atlas

A collection of cross-cancer analyses delving into overarching themes on cancer, including cell-of-origin patterns, oncogenic processes and signaling pathways. Published in 2018 at the

https://www.youtube.com/watch?time_continue=249&v=epsZjJ_A1y4

<https://cancergenome.nih.gov/>

TCGA: Overview

- Initiated in 2005
- A joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI).
- 27 participating Institutes in US and Canada.
- The overarching goal of TCGA is to improve our ability to diagnose, treat and prevent cancer, through the application of genome analysis technologies, including large-scale genome sequencing.
- The Cancer Genome Atlas Network have published more than 20 papers since the project began

(<https://tcga-data.nci.nih.gov/docs/publications/>)

NATIONAL CANCER INSTITUTE THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

TCGA produced over

2.5
PETABYTES
of data



TCGA data describes

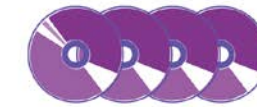
33
DIFFERENT
TUMOR TYPES

including

10
RARE
CANCERS

To put this into perspective, 1 petabyte of data is equal to

212,000
DVDs



...based on paired tumor and normal tissue sets collected from

11,000
PATIENTS

...using

7
DIFFERENT
DATA TYPES



TCGA RESULTS & FINDINGS



MOLECULAR
BASIS OF
CANCER

Improved our understanding of the genomic underpinnings of cancer



TUMOR
SUBTYPES

Revolutionized how cancer is classified



THERAPEUTIC
TARGETS

Identified genomic characteristics of tumors that can be targeted with currently available therapies or used to help with drug development

For example, a TCGA study found the basal-like subtype of breast cancer to be similar to the serous subtype of ovarian cancer on a molecular level, suggesting that despite arising from different tissues in the body, these subtypes may share a common path of development and respond to similar therapeutic strategies

TCGA revolutionized how cancer is classified by identifying tumor subtypes with distinct sets of genomic alterations.*

TCGA's identification of targetable genomic alterations in lung squamous cell carcinoma led to NCI's Lung-MAP Trial, which will treat patients based on the specific genomic changes in their tumor.

THE TEAM



20
COLLABORATING
INSTITUTIONS
across the United States
and Canada

WHAT'S NEXT?

The Genomic Data Commons (GDC) houses TCGA and other NCI-generated data sets for scientists to access from anywhere. The GDC also has many expanded capabilities that will allow researchers to answer more clinically relevant questions with increased ease.



*TCGA's analysis of stomach cancer revealed that it is not a single disease, but a disease composed of four subtypes, including a new subtype characterized by infection with Epstein-Barr virus.

TCGA Data Portal

<https://portal.gdc.cancer.gov/>

Harmonized Cancer Datasets

Genomic Data Commons Data Portal

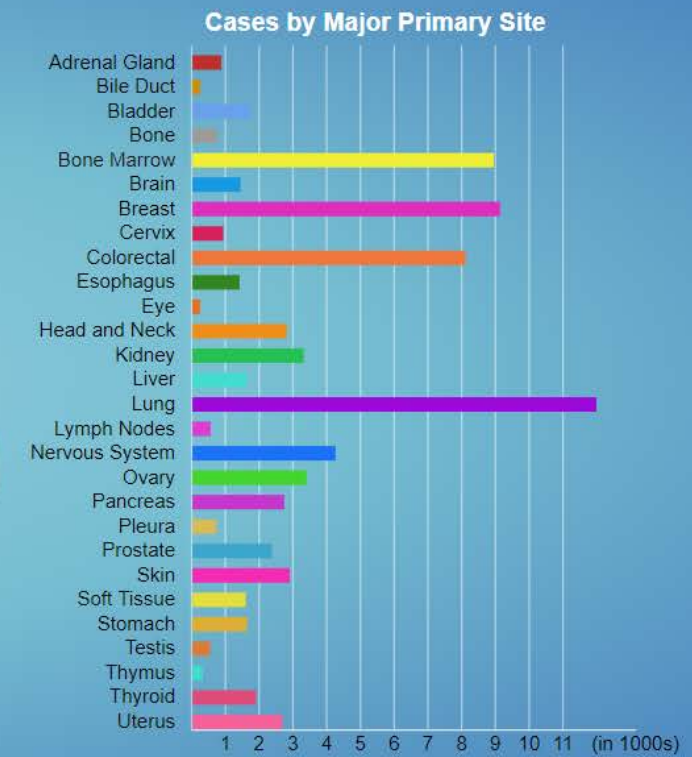
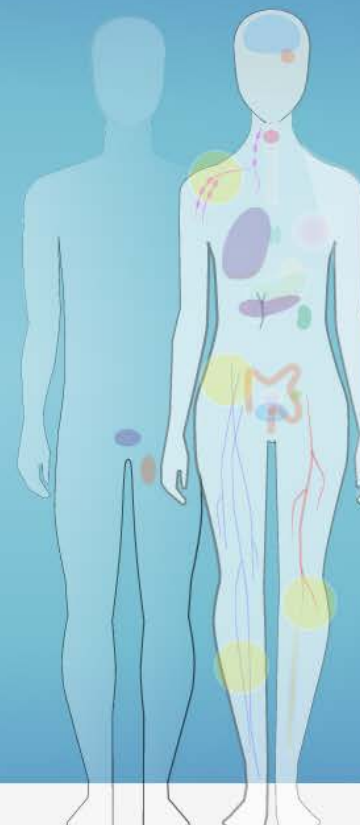
Get Started by Exploring:

- Projects
- Exploration
- Analysis
- Repository

Q e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary [Data Release 31.0 - October 29, 2021](#)

PROJECTS 70	PRIMARY SITES 67	CASES 85 415
FILES 649 152	GENES 23 621	MUTATIONS 3 599 319



GDC Applications

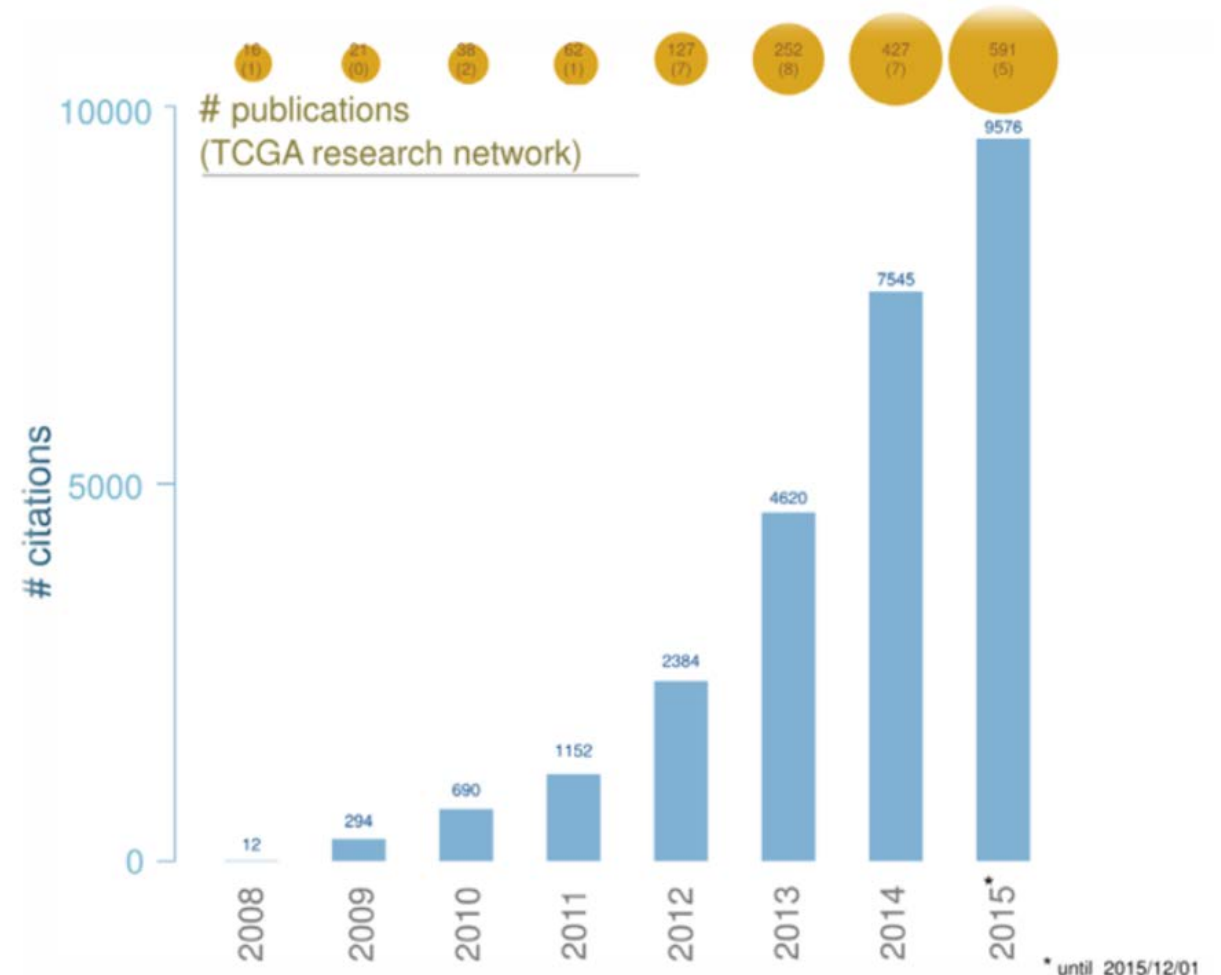
The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

TCGA: A Valuable Resource for Research Community

TCGA Data Types

- Clinical data
- DNA sequencing
- miRNA sequencing
- Protein expression
- mRNA sequencing
- Total RNA sequencing
- Array-based expression
- DNA methylation
- Copy number variations

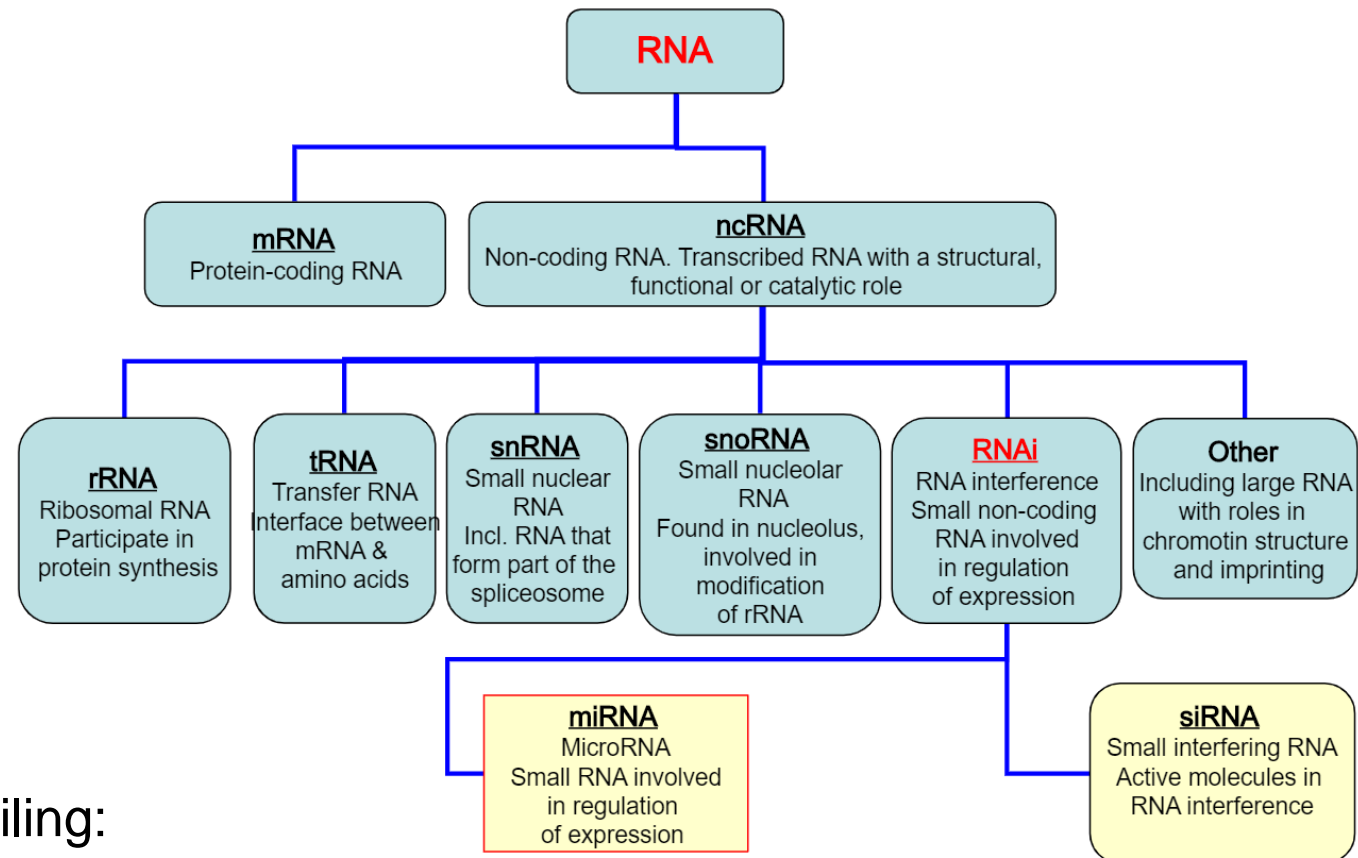
+ Computational tools



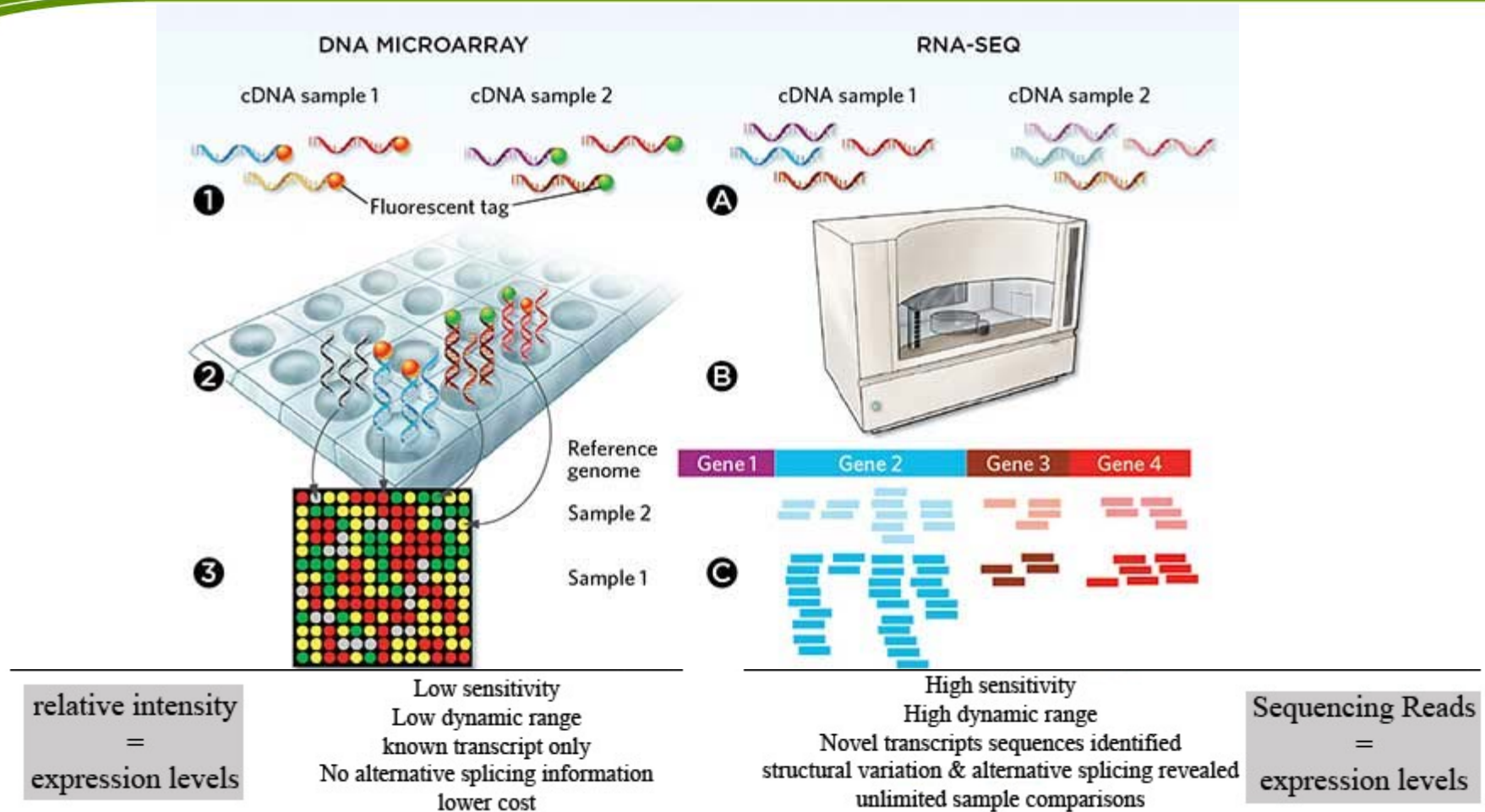
Transcriptomics

- Study of transcriptome, the sum of all RNA transcripts
- Two most widely studied types of RNA
 - mRNA - transcriptome or the expressed genes. Usually contains genes with poly A tail.
 - miRNA - Small non-coding RNA (containing about 21-25 nucleotides), important in gene regulation.
- Array-based Expression Profiling:
- <https://www.youtube.com/watch?v=6ZzFihESjp0>

Type of RNA molecules

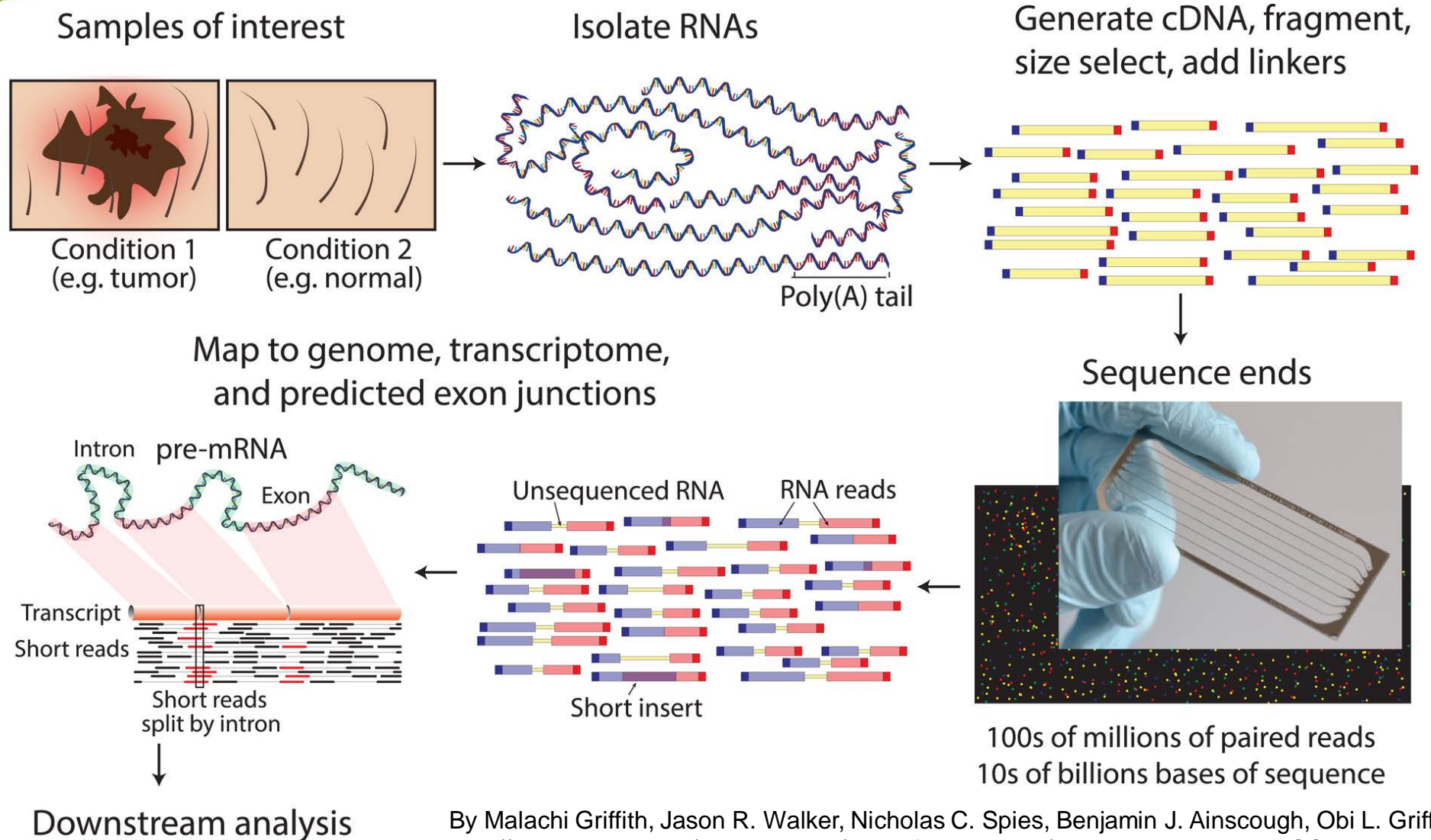


Microarrays vs RNA-seq



- While methods for analyzing microarray data are fully mature and straightforward, there is no consensus on which pipelines—or series of computational steps—to use to analyze RNA-seq data.

Overview of RNA-seq

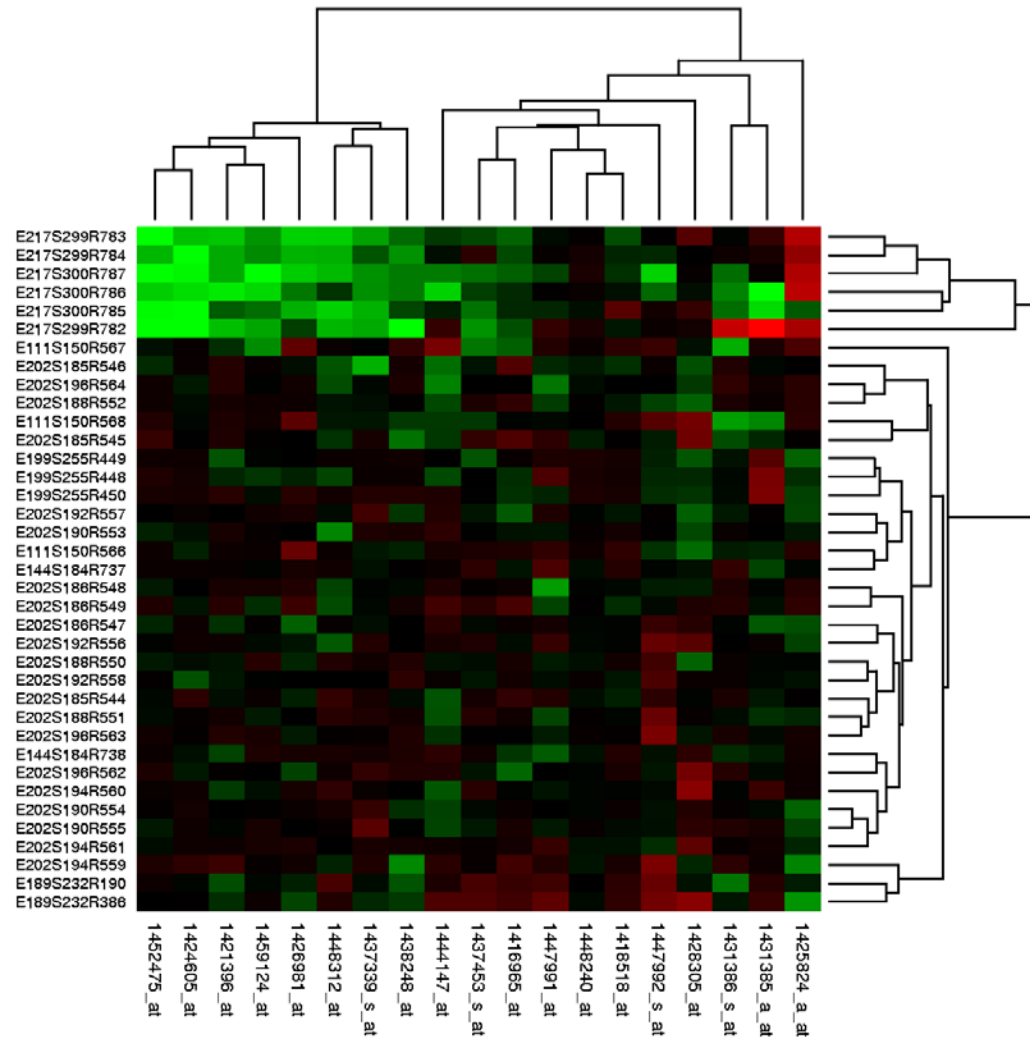


By Malachi Griffith, Jason R. Walker, Nicholas C. Spies, Benjamin J. Ainscough, Obi L. Griffith - <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004393>, CC BY 2.5, <https://commons.wikimedia.org/w/index.php?curid=53055894>

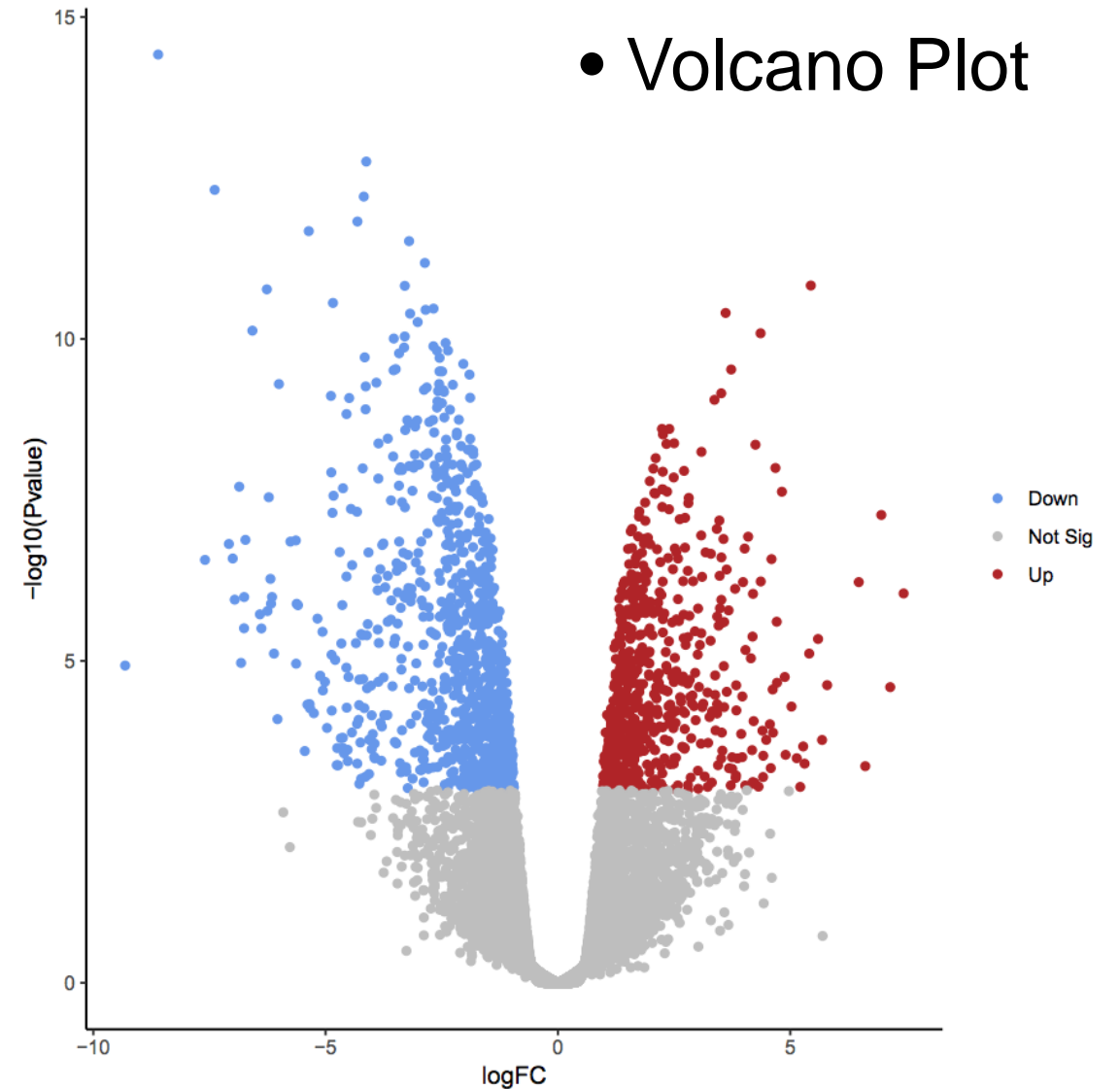
RNA sequencing downstream analysis

- <https://www.youtube.com/watch?v=tlf6wYJrwKY> (from 13:10)
- More info about microarray vs. RNA-seq at:
<https://www.youtube.com/watch?v=2c3t3tDEmsU>
- More info RNA seq at:
- https://www.youtube.com/watch?v=MFRkwXq6v_I
- Useful detailed info about anything connected to RNA-seq
- <https://www.rna-seqblog.com>

Examples of transcriptomics data outputs



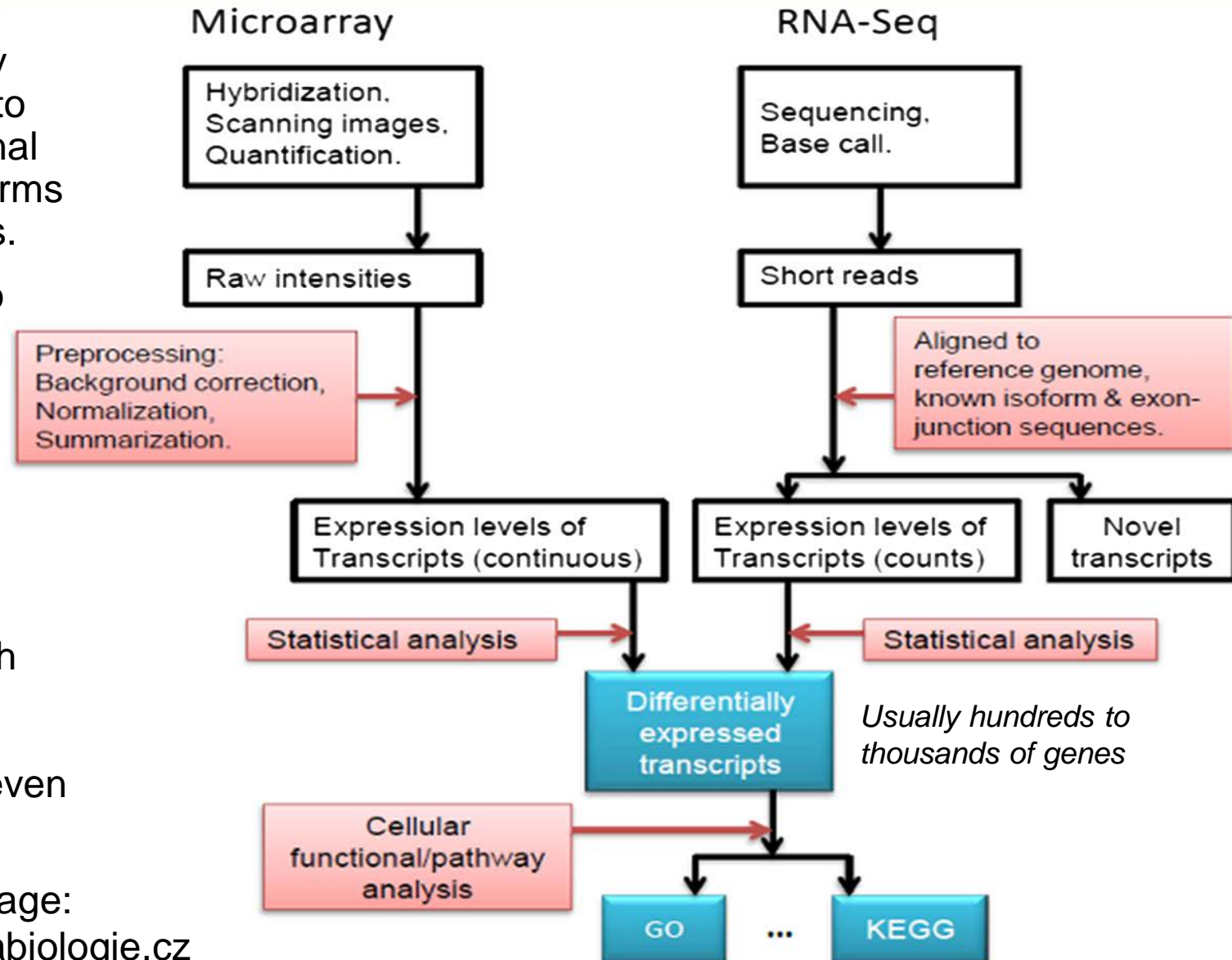
• Heat map



• Volcano Plot

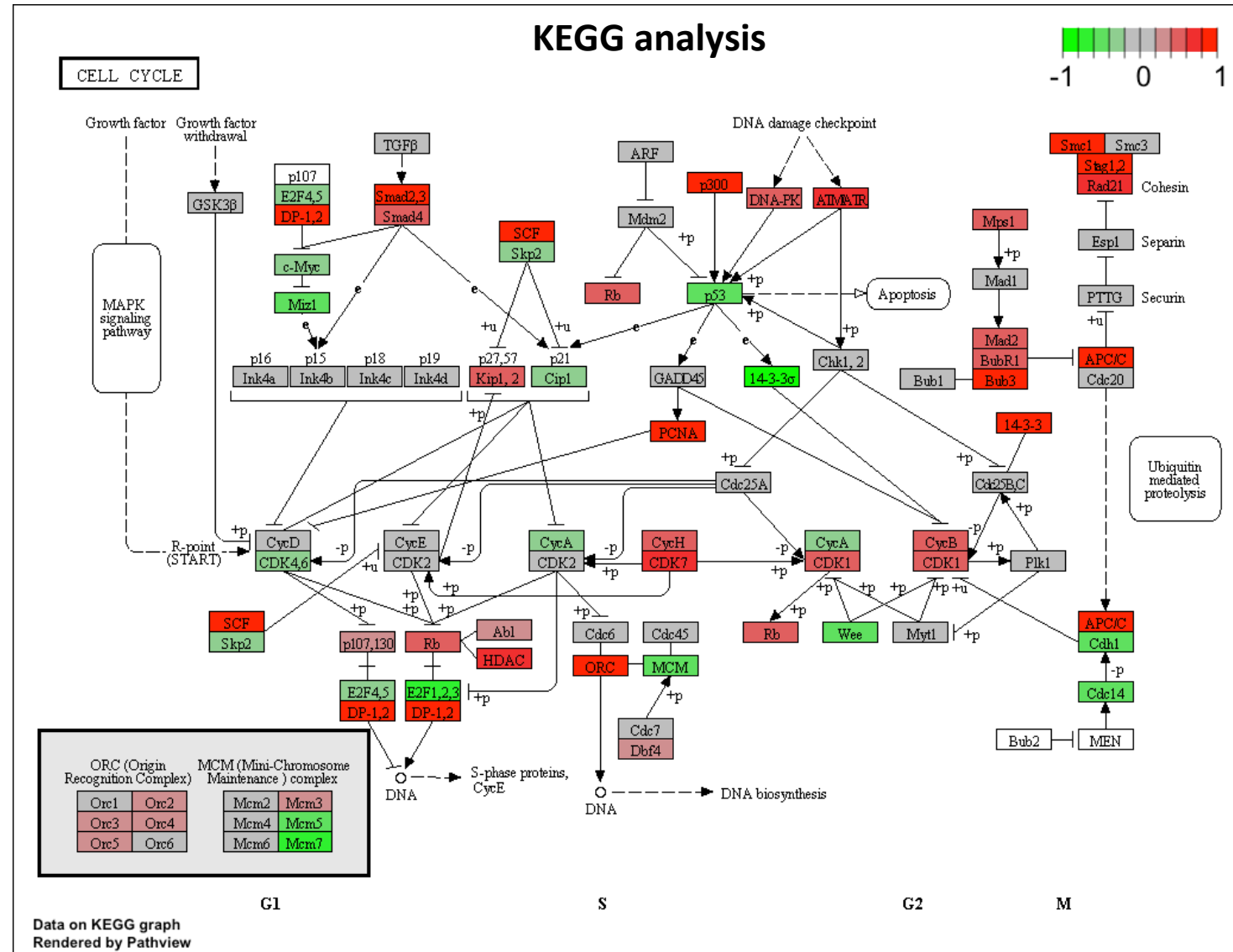
Cellular/functional/pathway analysis

- Cellular/functional/pathway analysis is a valuable tool to summarize high-dimensional gene expression data in terms of biologically relevant sets.
- Genes are aggregated into gene sets on the basis of shared biological or functional properties as defined by a reference knowledge base.
- Knowledge bases are database collections of molecular knowledge which may include molecular interactions, regulation, molecular product(s) and even phenotype associations.
- Useful info in Czech language: <https://portal.matematickabiologie.cz>



Database resources for understanding high-level functions and utilities of the biological system

- Database tools:
 - **KEGG** (Kyoto Encyclopedia of Genes and Genomes)
 - (<https://www.genome.jp/kegg/>)
 - Disadvantage – does not provide statistical significance of particular pathways
 - And many others available online



Gene-set analysis (GSA)/Pathway analysis



[About](#) [Ontology](#) [Annotations](#) [Downloads](#) [Help](#)

- Gene Ontology (GO) analysis (<http://geneontology.org/>)



Current release 2021-10-26: 43 832 GO terms | 7 827 476 annotations
1 542 582 gene products | 5 086 species ([see statistics](#))

THE GENE ONTOLOGY RESOURCE

The mission of the GO Consortium is to develop a comprehensive, **computational model of biological systems**, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life.

The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research.

Search GO term or Gene Product in AmiGO ...



Any Ontology Gene Product

GO Enrichment Analysis

Powered by PANTHER

Your gene IDs here...

biological process

Homo sapiens

Examples

Launch

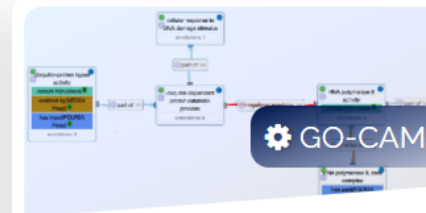
Hint: can use UniProt ID/AC, Gene Name, Gene Symbols, MOD IDs



The network of biological classes describing the current best representation of the "universe" of biology: the molecular functions, cellular locations, and processes gene products may carry out.



Statements, based on specific, traceable scientific evidence, asserting that a specific gene product is a real exemplar of a particular GO class.



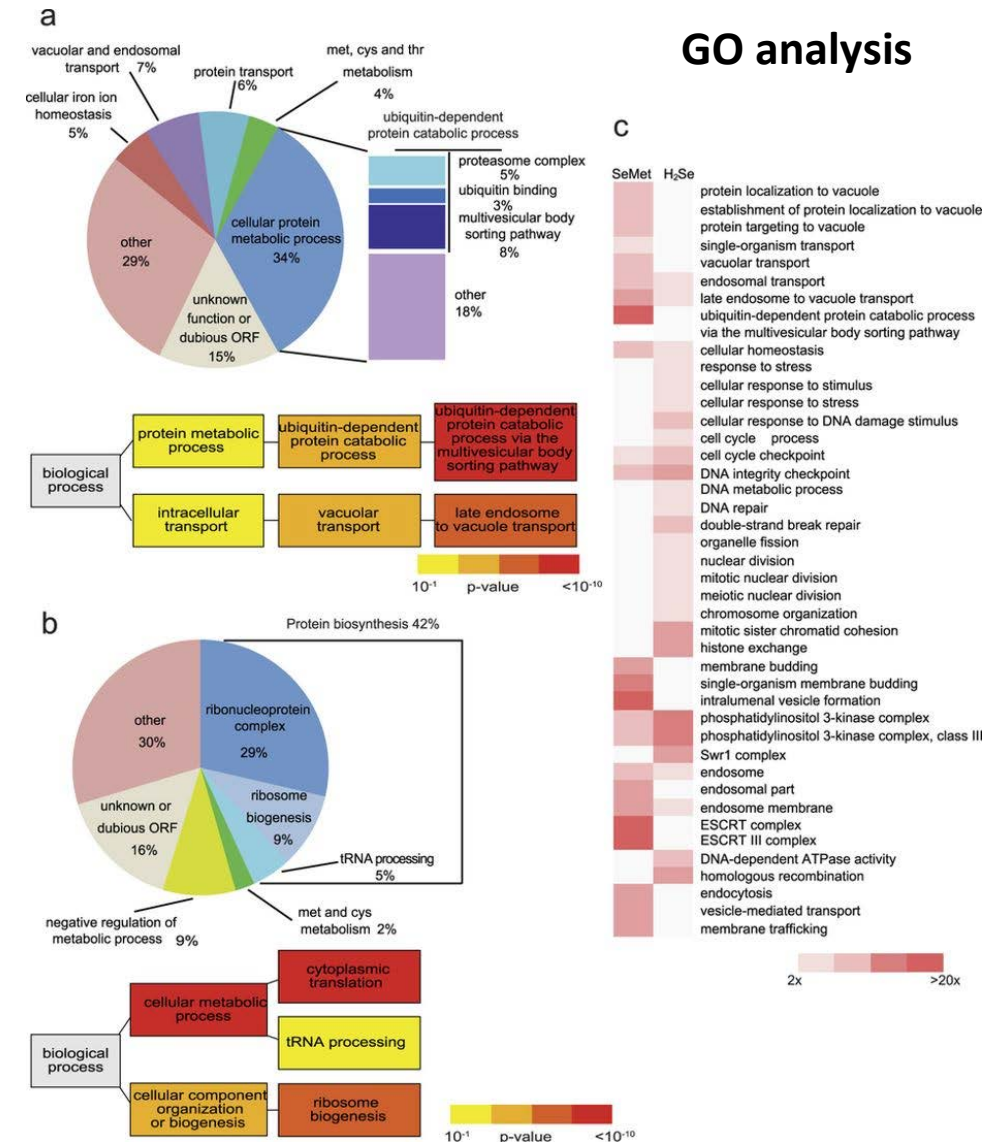
GO Causal Activity Model (GO-CAM) provides a structured framework to link standard GO annotations into a more complete model of a biological system.



Tools to curate, browse, search, visualize and download both the ontology and annotations. Includes bioinformatic guides (Notebooks) and simple API access to integrate the GO into your research.

Example data of GO enrichment analysis

- **GO enrichment analysis**
- One of the main uses of the GO is to perform enrichment analysis on gene sets. For example, given a set of genes that are up-regulated under certain conditions, an enrichment analysis will find which GO terms are over-represented (or under-represented) using annotations for that gene set.
- 3 main GO aspects (molecular function, biological process, cellular component)
- <http://geneontology.org/docs/go-enrichment-analysis/>



Reactome Knowledgebase

Why Reactome

Reactome is a free, open-source, curated and peer-reviewed pathway database. Our goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education.

If you use Reactome in Asia, we suggest using our Chinese mirror site at reactome.ncpsb.org.cn.

EMBL-EBI NYU Langone Health OICR

The development of Reactome is supported by grants from the US National Institutes of Health (U41 HG003751) and the European Molecular Biology Laboratory.

Latest News

We want to hear your Success Story!

- Version 78 Released
- Reactome Multi-Omics Pathway Analysis webinar reaches record attendance
- Version 77 Released
- The Reactome IDG Portal is released

Tweets

reactome @reactome
An interesting publication just out using Reactome analysis tools & textbook-style illustrations #citingreactome <https://twitter.com/guanwg/status/1456253300098035715>
4 Nov 2021


reactome @reactome
Introducing "Success Story of the Month"! Have you had some success with your experiment, tool or resource by using Reactome? Submit your #usecase success story, more details: reactome.org/about/news/172...

Success Story [View on Twitter](#)


- More info at:
- <https://www.youtube.com/user/Reactome/videos>

 Version 78 released on October 13, 2021



2,546
Human Pathways


13,890
Reactions


10,720
Proteins

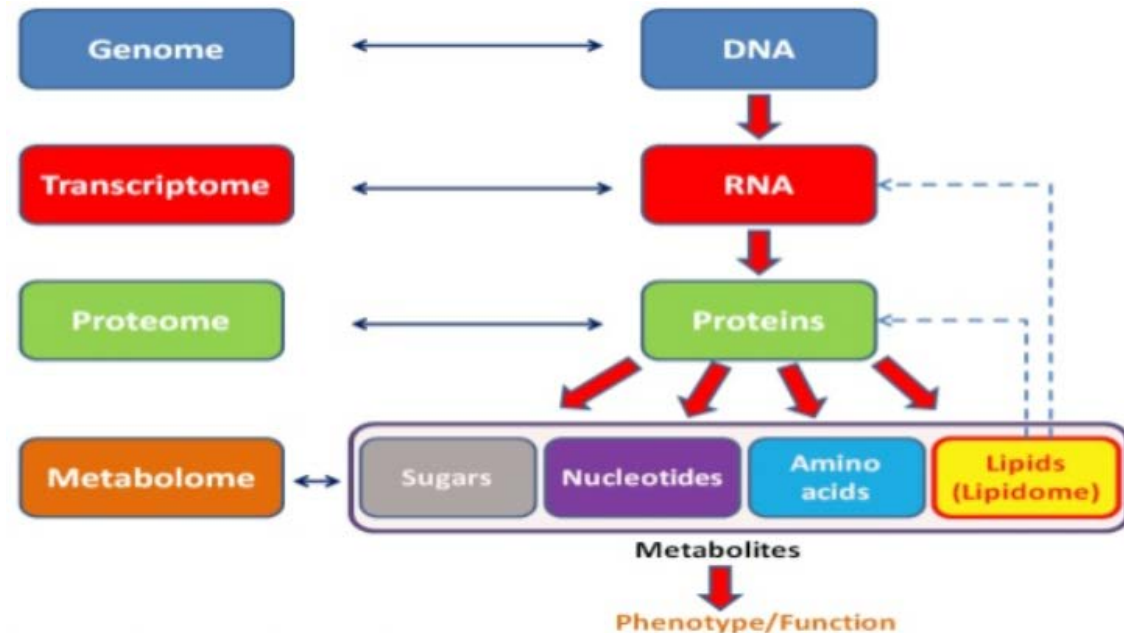
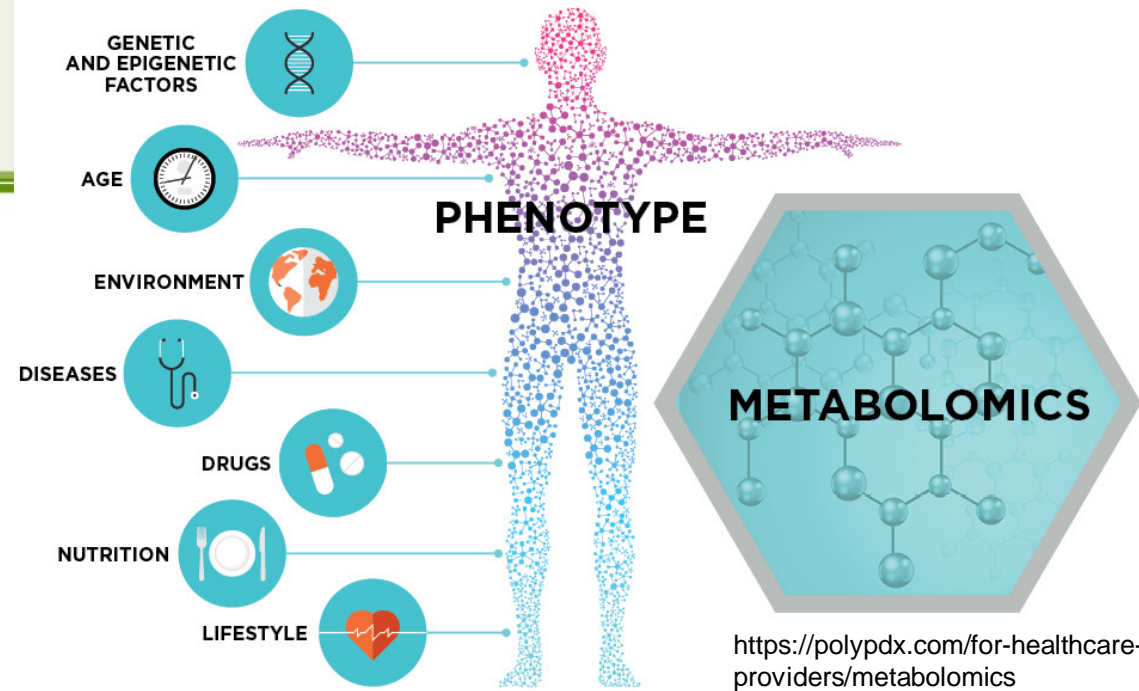

1,940
Small Molecules


507
Drugs


34,025
Literature References

Metabolomics

- Metabolomics – large-scale systematic study of the metabolome
- Metabolome - total complement of metabolites present in a biological sample under given genetic, nutritional or environmental conditions
 - the unique biochemical fingerprint of all cellular processes
- Metabolite - low molecular (usually 50 – 1,500 Da) weight organic compound, typically involved in a biological process as a substrate or product.
- Metabolomics yield many insights into basic biological research in areas such as systems biology, metabolic modelling, pharmaceutical research, nutrition and toxicology

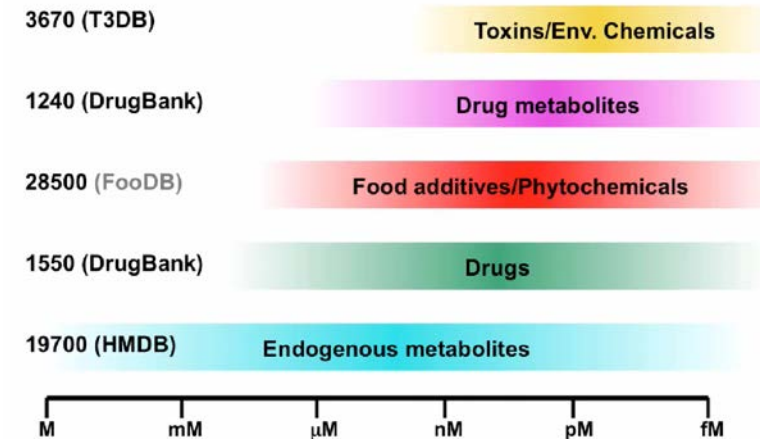


Metabolites are important

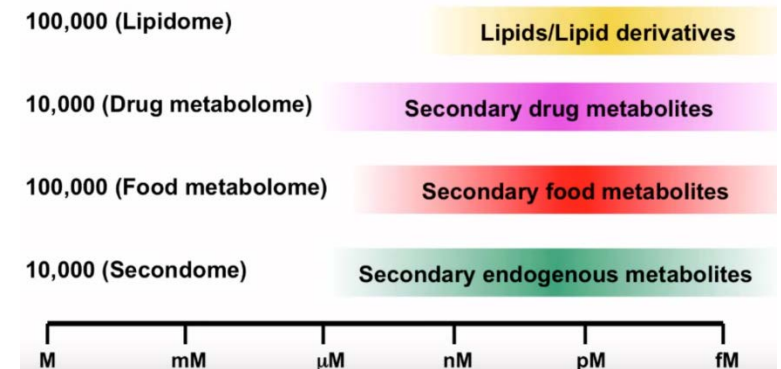
- **>95% of all diagnostic clinical assays test for small molecules**
- **89% of all known drugs are small molecules**
- **50% of all drugs are derived from pre-existing metabolites**
- **30% of identified genetic disorders involve diseases of small molecule metabolism**
- **Small molecules serve as cofactors and signaling molecules to 1000's of proteins**

Metabolomics can therefore be seen as bridging the gap between genotype and phenotype

Human Metabolomes (2015)



Theoretical Human Metabolomes

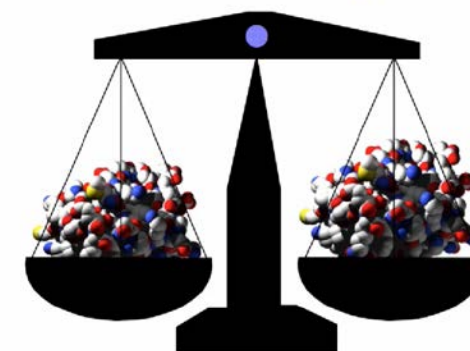


Metabolomics technologies

- UPLC, HPLC
- CE/microfluidics
- LC-MS
- FT-MS
- QqQ-MS
- NMR spectroscopy
- X-ray crystallography
- GC-MS
- FTIR

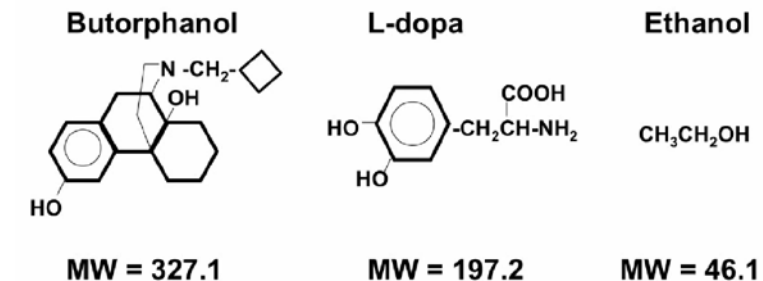
Mass Spectrometry

Analytical method to measure the molecular or atomic weight of samples



MS Principles

- Different compounds can be uniquely identified by their mass



Metabolomics – ,a snapshot‘ in time

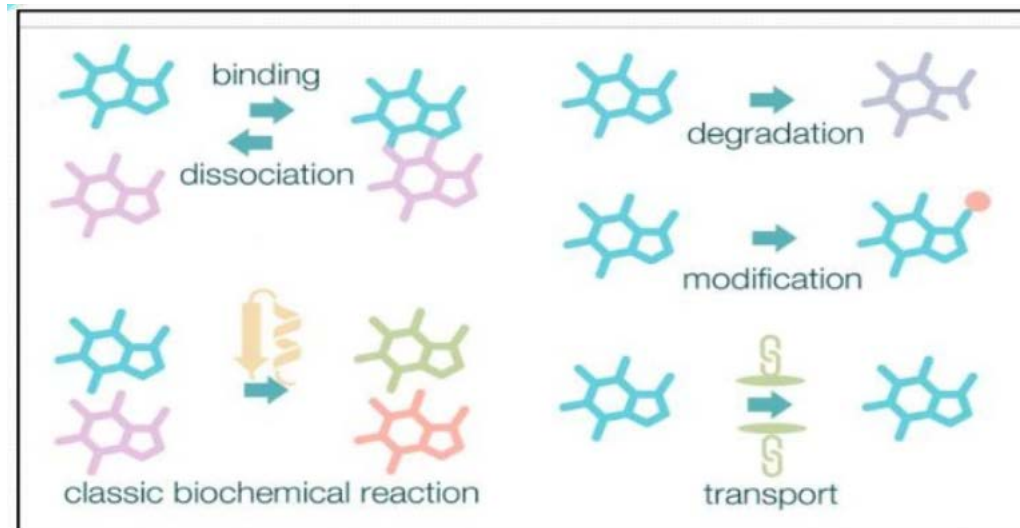
Conceptual approaches in metabolomics:

- Target analysis: has been applied for many decades and includes the determination and quantification of a small set of known metabolites (targets) using one particular analytical technique of best performance for the compounds of interest.

- Metabolite profiling: aims at the analysis of a larger set of compounds, both identified and unknown with respect to their chemical nature. This approach has been applied for many different biological systems using GC-MS, including plants, microbes, urine, and plasma samples.

- Metabolomics: employs complementary analytical methodologies, for example, LC-MS/MS, GC-MS, and/or NMR, in order to determine and quantify as many metabolites as possible, either identified or unknown compounds.

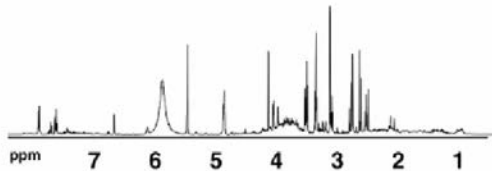
- Metabolic fingerprinting: a metabolic “signature” or mass profile of the sample of interest is generated and then compared in a large sample population to screen for differences between the samples. When signals that can significantly discriminate between samples are detected, the metabolites are identified and the biological relevance of that compound can be elucidated, greatly reducing the analysis time.



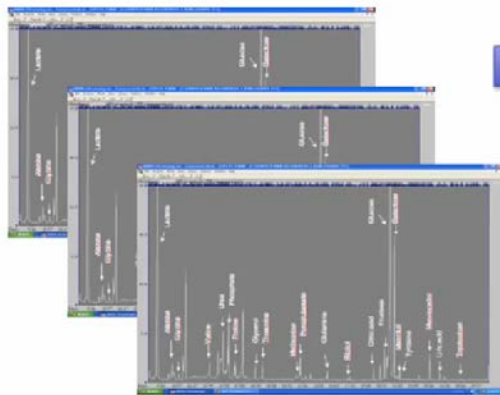
A diagram showing the main different types of metabolic reactions that take place in a cell. These are shown as they are represented in the database *Reactome*.

Metabolomics data analysis

From Spectra to Lists



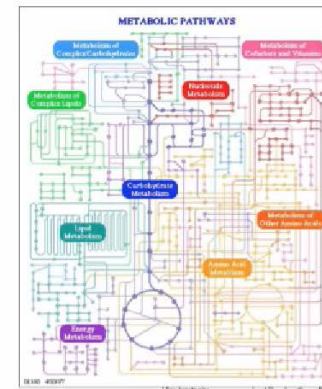
From Lists to Pathways



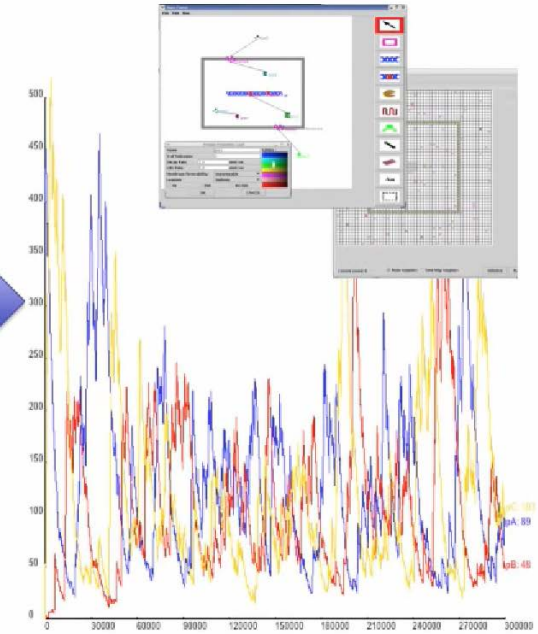
Compound	Retention Time (min)	Conc. in Urine (µM)	Compound	Retention Time (min)	Conc. in Urine (µM)
Dns-o-phospho-L-leucine	0.92	<0.L.*	Dns-ile	6.35	25
Dns-o-phospho-L-tyrosine	0.95	<0.L.	Dns-3-aminosalicylic acid	6.44	0.5
Dns-adenosine monophosphate	9.09	<0.L.	Dns-pipecolic acid	6.50	0.5
Dns-o-phosphoethanolamine	1.06	96	Dns-4-ure	6.54	54
Dns-glucosamine	1.06	22	Dns-cystathionine	6.54	0.3
Dns-o-phospho-L-threonine	1.09	<0.L.	Dns-L-iso-Pro	6.60	0.4
Dns-6-dimethyl-histamine putres	1.30	<0.L.	Dns-5-hydroxylysine	6.65	1.6
Dns-3-methyl-histidine	1.22	80	Dns-Cytidine	6.73	100
Dns-leucine	1.25	834	Dns-N-norleucine	6.81	0.1
Dns-carnitine	1.34	28	Dns-5-hydroxyproline	7.17	<0.L.
Dns-Arg	1.53	36	Dns-dimethylamine	7.33	205
Dns-Asn	1.55	133	Dns-5-HIAA	7.46	18
Dns-hypoxanthine	1.58	10	Dns-sambaliferone	7.47	1.9
Dns-homocysteine	1.61	3.9	Dns-2,3-diaminopropionic acid	7.53	<0.L.
Dns-guanidine	1.62	<0.L.	Dns-L-ornithine	7.70	15
Dns-Gln	1.72	633	Dns-4-acetylmorphine	7.73	51
Dns-allantoin	1.83	3.8	Dns-proline	7.73	8.9
Dns-L-cystathionine	1.87	2.9	Dns-8-homocysteine	7.76	3.3
Dns-1-(or 3)-methylhistamine	1.94	1.9	Dns-acetaminophen	7.97	82
Dns-adenosine	2.06	2.6	Dns-Phe-Phe	8.03	0.4
Dns-methylglycine	2.20	<0.L.	Dns-5-methyl-xytalcolic acid	8.04	3.1
Dns-Ser	2.24	511	Dns-L-lys	8.16	194
Dns-aspartic acid amide	2.44	26	Dns-argin	8.17	<0.L.
Dns-4-hydroxy-proline	2.56	2.3	Dns-iso-Phe	8.22	0.3
Dns-Orn	2.57	21	Dns-ile	8.35	1860
Dns-Asp	2.60	90	Dns-4-thalidomide	8.37	<0.L.
Dns-Thr	3.03	107	Dns-benzylamine	8.38	<0.L.
Dns-epinephrine	3.05	<0.L.	Dns-1-nephedrine	8.50	0.6
Dns-ethanolamine	3.11	471	Dns-tryptamine	8.53	0.4
Dns-aminoadipic acid	3.17	80	Dns-pyridoxamine	8.94	<0.L.
Dns-Gly	3.43	2510	Dns-2-methyl-phenylethylamine	9.24	<0.L.
Dns-ala	3.58	658	Dns-6-hydroxytryptophan	9.25	0.12
Dns-aminolevulinic acid	3.97	30	Dns-1,3-diaminopropane	9.44	0.25
Dns-ε-amino-4-tyric acid	3.98	4.6	Dns-putrescine	9.60	0.5
Dns-γ-aminol-hippuric acid	3.98	2.9	Dns-1,2-diaminopropane	9.66	0.1
Dns-5-hydroxy-methyluracil	4.58	1.9	Dns-lysosaminamide	9.75	29
Dns-4-tyrosylasparagine	4.70	5.5	Dns-dopamine	10.06	140
Dns-tyrosine	4.75	<0.L.	Dns-squalonine	10.06	0.06
Dns-5-aminopentanoic acid	4.79	1.6	Dns-β-alanine	10.19	0.4
Dns-ascorbic acid	4.81	7.2	Dns-3-methoxy-tyramine	10.19	9.2
Dns-3-amino-isobutyrate	4.81	85	Dns-Thr	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns-cysteamine	10.41	<0.L.



From Pathways & Lists to Models & Biomarkers



Compound	Retention Time (min)	Conc. in Urine (µM)
Dns-tyrosine	10.19	9.2
Dns-tryptamine	8.53	0.4
Dns-tryptophan	10.19	9.2
Dns-tryptophan-5-hydroxytryptophan	10.19	9.2
Dns-tryptophan-5-hydroxytryptophan-5-hydroxytryptophan	10.19	9.2
Dns-tryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan	10.19	9.2
Dns-tryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan	10.19	9.2
Dns-tryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan	10.19	9.2
Dns-tryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan	10.19	9.2
Dns-tryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan	10.19	9.2
Dns-tryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan	10.19	9.2
Dns-tryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan-5-hydroxytryptophan	10.19	9.2



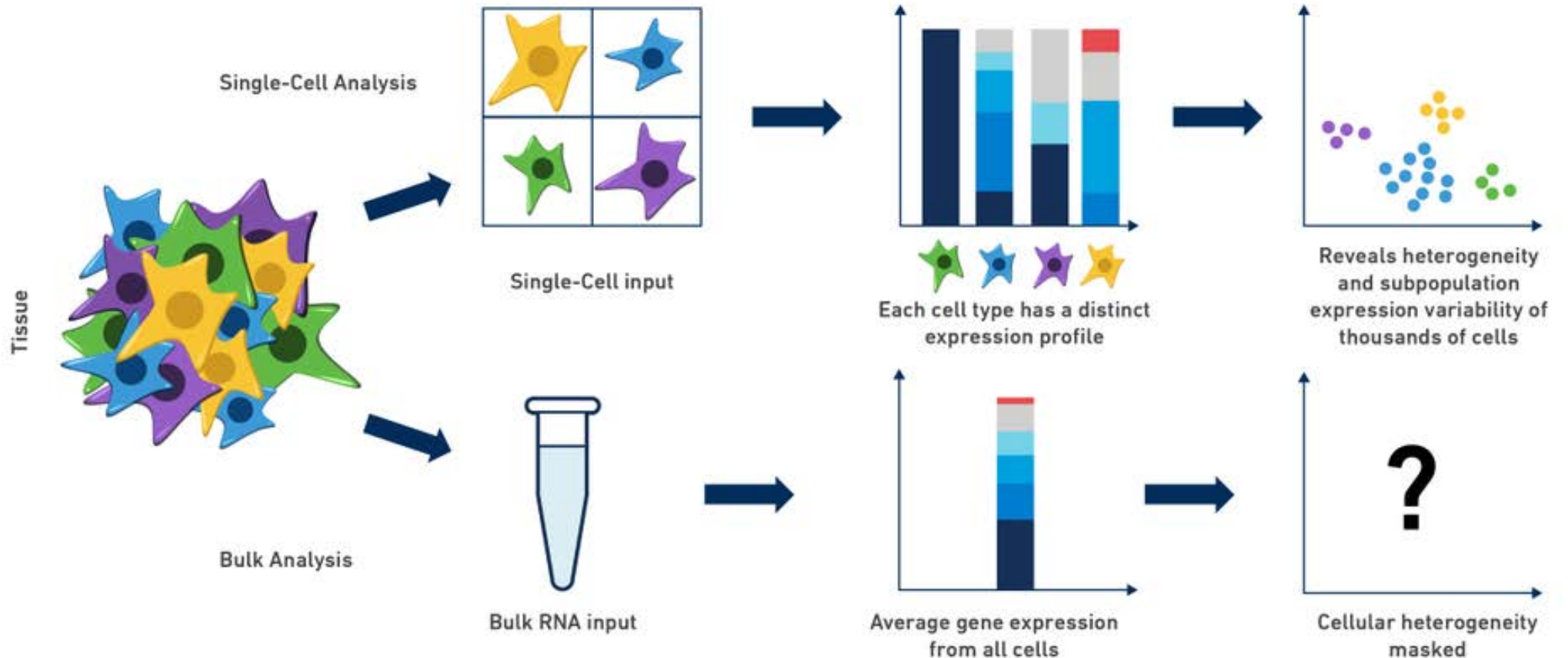
Where to look for metabolomics data

Metabolic pathway databases

- ▶ Pathway viewers KEGG (<http://www.genome.ad.jp/kegg/>),
- ▶ Atomic Reconstruction of Metabolism database (<http://www.metabolome.jp/>),
- ▶ BioCyc (<http://biocyc.org>) (Paley and Karp 2006),
- ▶ MetaCyc (<http://metacyc.org/>) (Caspi et al. 2006),
- ▶ AraCyc (<http://www.Arabidopsis.org/tools/aracyc/>) (Zhang et al. 2005), MapMan (<http://gabi.rzpd.de/projects/MapMan/>)
- ▶ (Thimm et al. 2004), KaPPA-View (<http://kpv.kazusa.or.jp/kappa-view/>) (Tokimatsu et al. 2005) and
- ▶ BioPathAT (<http://www.ibr.wsu.edu/research/lange/public%5Ffolder/>) (Lange and Ghassemian 2005),
- ▶ the data model for plant metabolomics experiments ArMet (<http://www.armet.org/>)

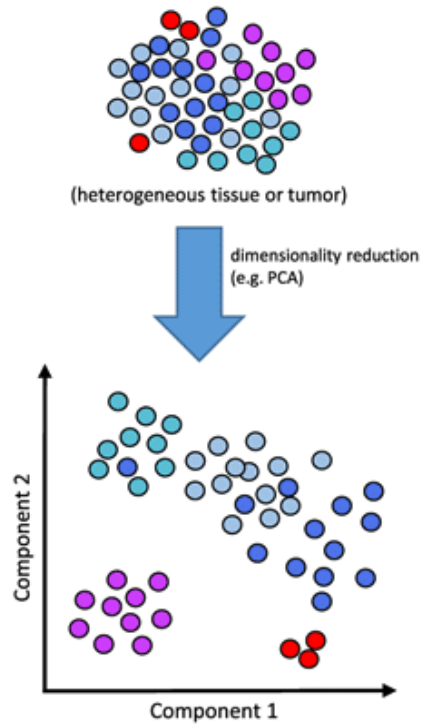
Cutting edge: Single-cell -omics

- Application of whole genome, whole transcriptome sequencing and other -omics methods to single cells, scRNA-seq is now the top method

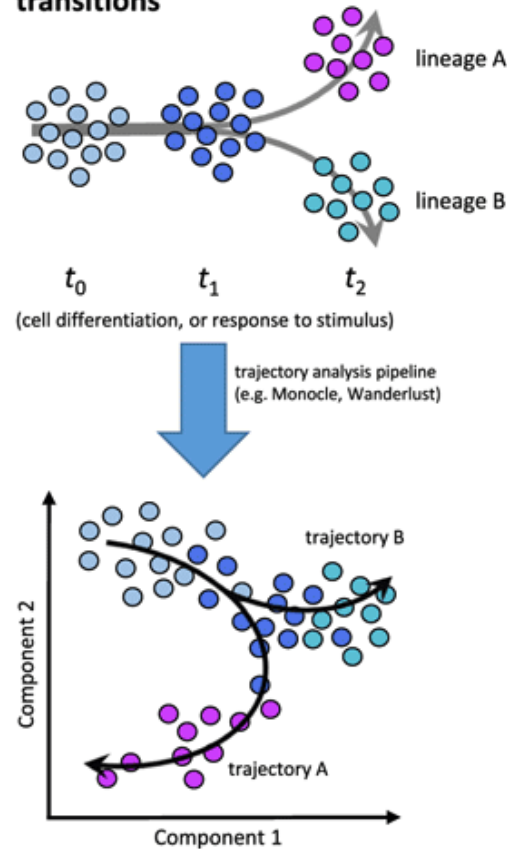


Common applications of scRNA-seq

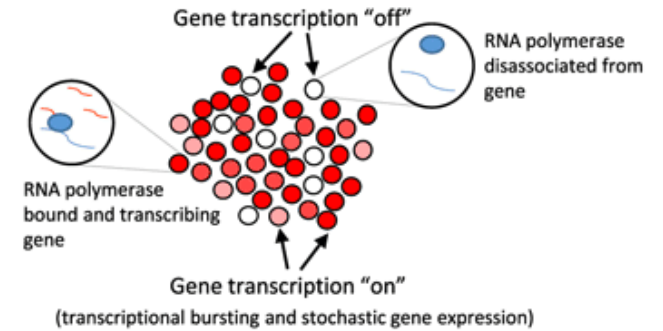
a) Deconvolving heterogeneous cell populations



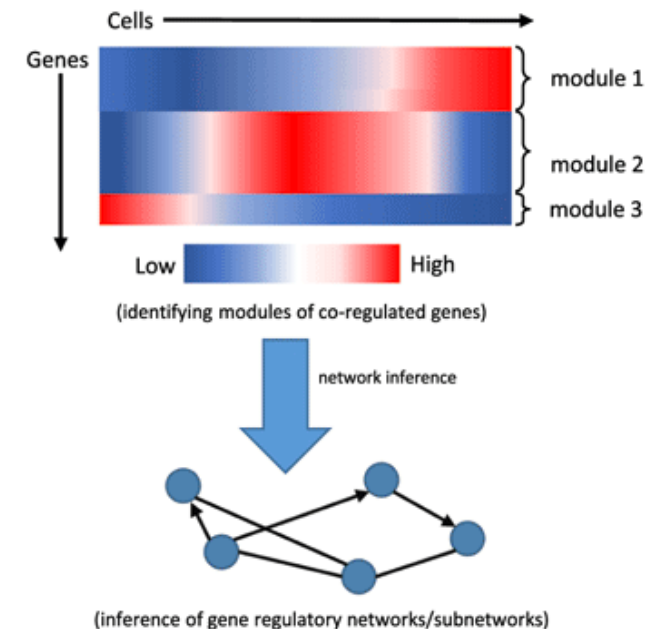
b) Trajectory analysis of cell state transitions



c) Dissecting transcription mechanics



d) Network inference



<https://f1000research.com/articles/5-182/v1>

For more info go at: <https://omicstools.com>

ScRNA-seq databases


<https://www.ebi.ac.uk/gxa/sc/home>







EMBL-EBI Services Research Training About us

Single Cell Expression Atlas

Single cell gene expression across species

Query bulk expression
[Back to Expression Atlas](#)



 Homo sapiens 1518 experiments Baseline: 79 Differential: 1439	 Mus musculus 1185 experiments Baseline: 49 Differential: 1136	 Rattus norvegicus 152 experiments Baseline: 3 Differential: 149	 Drosophila melanogaster 142 experiments Baseline: 4 Differential: 138	 Gallus gallus 39 experiments Baseline: 4 Differential: 35	 Caenorhabditis elegans 30 experiments Baseline: 1 Differential: 29
--	--	--	--	--	---

Home Gene search Browse experiments Release notes Help Support

Search across 18 species, 229 studies, 5 978 348 cells

Ensembl 104, Ensembl Genomes 51, WormBase ParaSite 15, EFO 3.10.0







Search

Gene ID or gene symbol

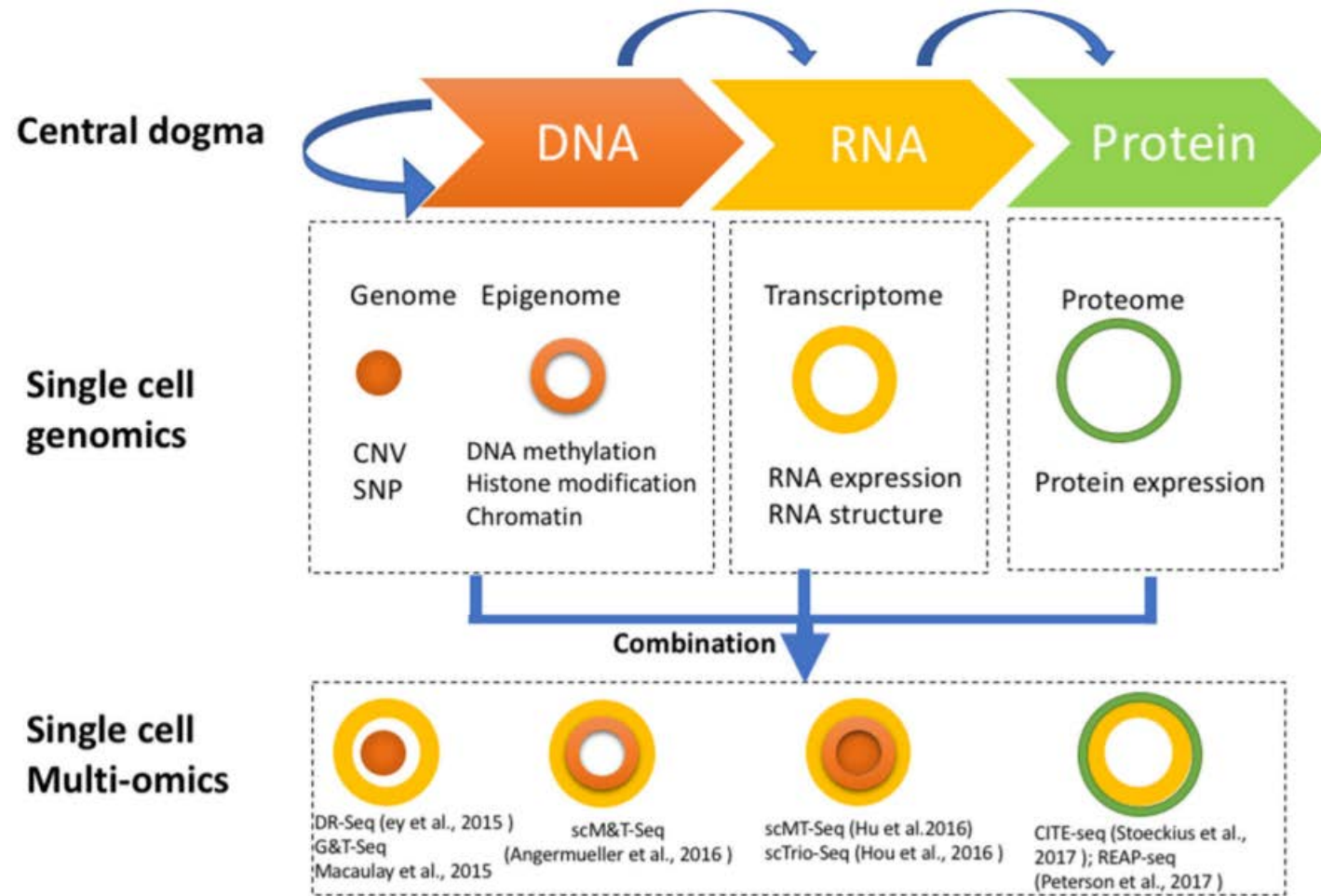
Species

Examples: [CFTR](#) (gene symbol), [ENSG00000115904](#) (Ensembl ID), [657](#) (Entrez ID), [MGI:98354](#) (MGI ID), [FBgn0004647](#) (FlyBase ID)

Animals Plants Fungi Protists

 Homo sapiens 103 experiments	 Mus musculus 76 experiments	 Drosophila melanogaster 9 experiments	 Danio rerio 7 experiments	 Gallus gallus 4 experiments	 Schistosoma mansoni 2 experiments
---	--	--	--	--	--

Single-cell multi-omics



Challenges:

- There are no commercial kits available yet for any single-cell multi-omics techniques, and many are technically challenging.
- Researchers must modify existing single-cell protocols so that they're compatible with multiple types of molecules and take great care to minimize the loss or contamination of samples
<https://www.the-scientist.com/lab-tools/integrating-multiple--omics-in-individual-cells-64829>

Difficulty squared

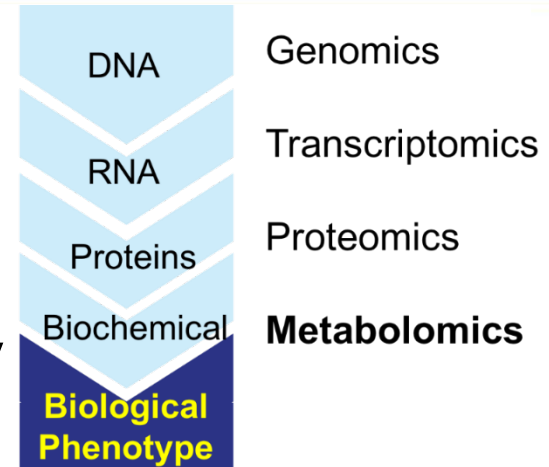
Combining modalities only multiplies the difficulty. All the weaknesses, all the noise, all the challenges from each technology, it just gets exacerbated by combining them into a multimodal assay.

Single-cell analysis enters the multiomics age
<https://www.nature.com/articles/d41586-021-01994-w#correction-1>

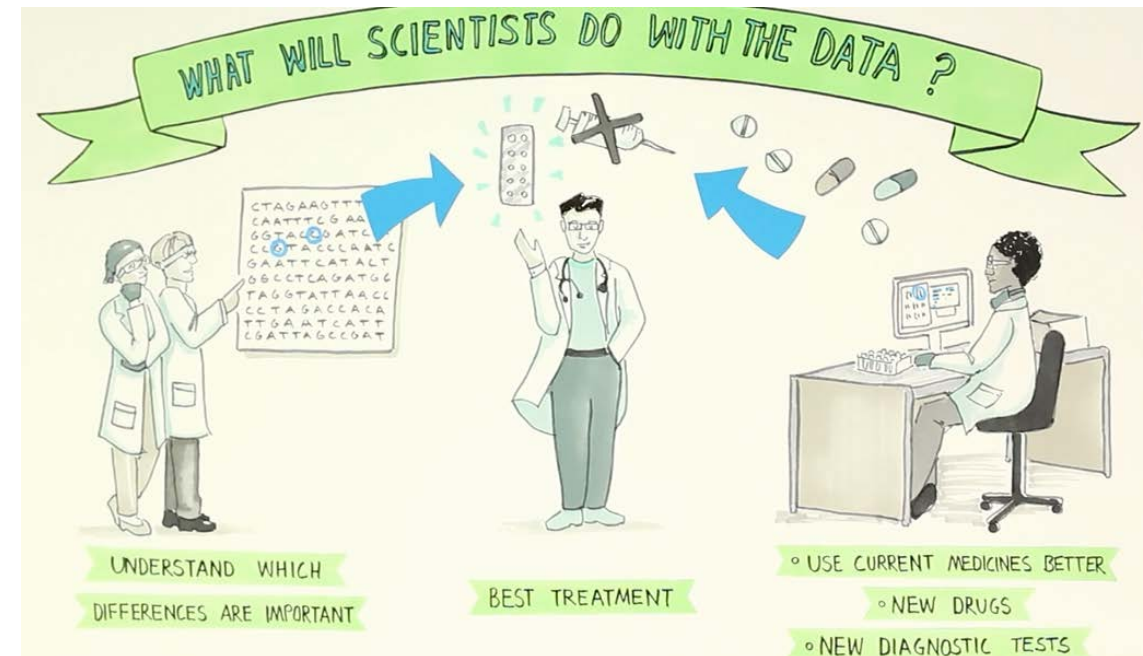
FIGURE 2 | Strategies for multi-omics profiling of single cells. Three major types of molecules relating to biological central dogma (Top). Single cell genomics methods profiling the genome, epigenome, transcriptome, and proteome are shown by different shapes with variable colors (Middle). Single cell multi-omics methods are built by combining different single cell sequencing methods to simultaneously profile multiple types of molecules of a single cell genome wide (Bottom). For example, G&T-seq was built by combining genome (orange) and transcriptome (yellow) to simultaneously detect DNA and RNA of the same cell genome wide.

Summary

- Omics technologies - „the data deluge“
- Genomics and Transcriptomics rely on two main approaches: microarrays (hybridization) and NGS (sequencing by synthesis)
- Proteomics and Metabolomics rely heavily on mass spectrometry



- Omics technologies are revolutionizing science and medicine
- From data to actionable knowledge - Integrated Omics data
- Precision medicine is the ultimate goal of many –omics efforts
- Despite the progress made we have still a long way to go ...



Take home messages

- We have been generating Big data, but we hardly understand it 😞
- Big data is publicly available, go through the databases before you even start planning your experiment – it can save you enormous time and money
- Databases contain huge datasets of patients you would never be able to gather by yourself, test your hypothesis in silico before the „wet-lab“ work
- If you cannot find the „yes/no“ or „a few genes“ answer, use the Cellular/functional/pathway analyses to help you out 😊
- Learning bioinformatics skills (e.g. programming in R) is a good investment plan for your future (scientific) career

Thank you for your attention

Any Questions?

- Jay Flatley, Executive Chairman of Illumina:
- *„Everyone is going to get sequenced, it is gonna be part of their health record and it will be used to manage their health care throughout their lifetime“.*

