

# PIRD 说明书

Beta 版

王珑龙(LONGLONG WANG)

# 目录

一、简介.....	1
二、数据库登录.....	1
三、主页.....	2
四、数据总览.....	4
4.1. 数据库展示形式.....	4
4.2. 数据展示页面的操作提示.....	4
页面内搜索.....	4
数据排序展示.....	5
导出选中数据.....	6
4.3. 项目汇总.....	6
4.4. 项目信息.....	7
4.5. 个体信息.....	8
4.6. 样本信息.....	9
4.6.1. 样本配置信息.....	9
4.6.2. 样本实验信息.....	10
4.6.3. 样本表型信息.....	10
4.6.4. 样本描述性统计信息.....	11
4.6.5. 项目全部样本概况汇总展示.....	12
4.6.6. 样本详情.....	13
五、检索查询.....	17
5.1. 检索项目.....	17
5.2. 检索样本.....	18
5.3. 检索克隆.....	18
5.4. BLAST 检索说明.....	19
检索示例.....	21
六、在线分析.....	27
6.1. 数据源.....	27
6.1.1. 用户自己上传的数据。.....	27

6.1.2. 数据库现有的公共数据。 .....	28
6.2. 数据表 .....	29
6.3. 可视化 .....	30
绘制简单的柱状图 .....	30
折线图 .....	32
箱线图 .....	33
七、知识库 .....	34
附录一、数据获取方法 .....	35

# PIRD 说明书 Beta 版

## 一、简介

免疫数据库 (PIRD) 收集了具有不同表型的人类和其他脊椎动物物种的免疫球蛋白 (IG) 和 T 细胞受体 (TCR) 的原始和分析的序列。您可以查看数据库中每个样本的详细信息, 根据需要选择样本进行分析, 并上传数据进行分析。您的分析结果将自动保存, 因此您可以随时返回查看。

## 二、数据库登录

中国国家基因库 (CNGB) 中有统一的用户注册及登录平台, 在该平台注册的账户可以用于登录 CNGB 的所有数据库。用户可以通过访问 <http://db.cngb.org/pird/> 网址进入 PIRD 数据库, 在您使用前请先点击主页顶端的用户帮助按钮并仔细阅读, 然后点击界面右上方的“登录/注册”按钮进入注册及登录页面。



## PIRD: Pan immune repertoire database

Pan immune repertoire database (PIRD) collects raw and processed sequences of immunoglobulins (IGs) and T cell receptors (TCRs) of human and other vertebrate species with different phenotypes. You can check the detailed information of each sample in the database, choose samples to analyze according to your need, and upload data to analyze. Your analysis results will be auto-saved, so you can return to check them at any time.

### Filter



Note: you can enter *status info, project name, species name, or article title* as keyword.

### Biological diversity

[10KP](#)[1KITE](#)[B10k](#)[Fish T1K](#)[Millet](#)[MLGD](#)[MT10K](#)[1KP](#)

### Health & Disease

[BDDB](#)[DHGV](#)

Language

## Login

[Log in](#)[Sign up an account](#)[Forget password?](#)

Log in with:

## 三、主页

数据库主页展示的内容包括数据库简介、检索查询栏、数据库的应用特点以及项目统计。用户可以通过主页的标签跳转到需要查看的功能页面。主页右侧提供了其他数据库及一些功能

模块的链接。



首页 项目 样本 检索查询 知识库 分析 敬请期待

## PIRD: 免疫数据库

免疫数据库 (PIRD) 收集了具有不同表型的人类和其他脊椎动物物种的免疫球蛋白 (IG) 和T细胞受体 (TCR) 的原始和分析的序列。您可以查看数据库中每个样本的详细信息, 根据需要选择样本进行分析, 并上传数据进行分析。您的分析结果将自动保存, 因此您可以随时返回查看。

### 检索查询

在这里输入您的关键字。



提示: 可以输入 健康状态, 项目名称, 物种名称, 或 样本编号 等字段的英文关键字。示例: health, breast, human.

### 应用特点

#### 检索信息

在PIRD中, 您可以检查每个样本的实验信息, 表型信息, 序列信息。

#### 创建项目

选择一些样品后, 您可以创建一个项目并保存分析结果。

#### 数据分析

我们将提供一些分析工具, 您可以使用这些工具进行数据分析。

#### 数据可视化

在PIRD中, 我们提供一些简洁优雅的数据可视化工具, 您可以使用这些工具轻松绘制一些高质量的图像。

### 项目统计



### 生物多样性



### 健康与疾病

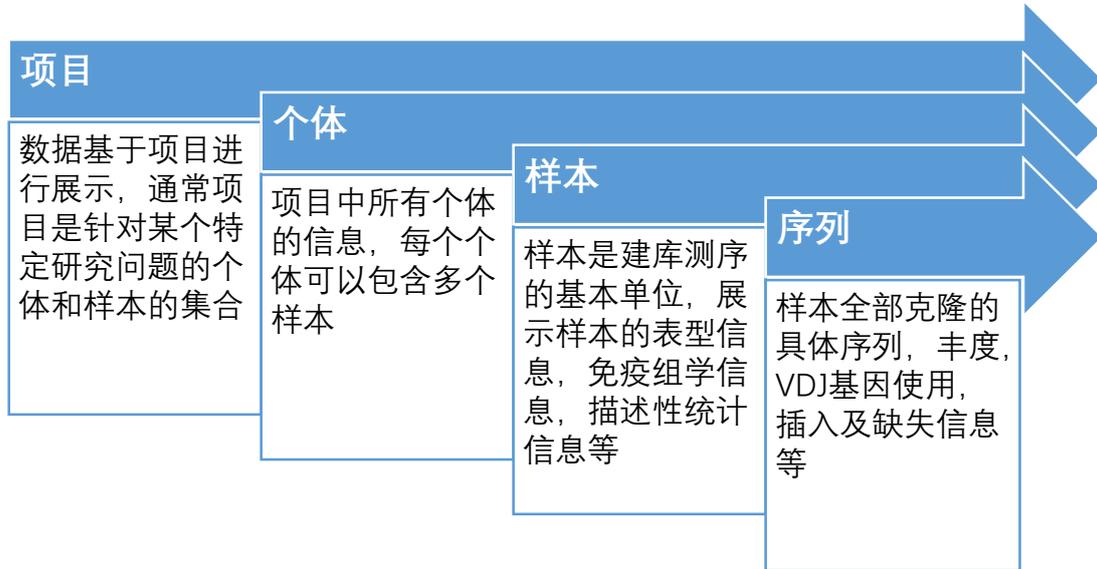


### 模块服务



## 四、数据总览

### 4.1. 数据库展示形式



### 4.2. 数据展示页面的操作提示

#### 页面内搜索

数据表格右上方有搜索栏，可以根据提示的字段进行搜索。

**PIRD**

首页 项目 样本 检索查询 知识库 分析 敬请期待

所有样本

样本配置信息

配置信息 实验信息 表型信息 描述性统计

导出选中数据 显示 10 页结果 搜索:

选择	项目	个体	样本	序列类型	物种	分子序列
<input type="checkbox"/>	colorectal_cancer	III11815201711110001	IS11815201710240673	RNA	human	IGH
<input type="checkbox"/>	colorectal_cancer	III11815201711110001	IS11815201710240674	RNA	human	IGH
<input type="checkbox"/>	colorectal_cancer	III11815201711110002	IS11815201710240675	RNA	human	IGH
<input type="checkbox"/>	colorectal_cancer	III11815201711110002	IS11815201710240676	RNA	human	IGH

可检索字段: 样本, 序列类型, 物种, 分子序列

# 数据排序展示



首页 项目 样本 检索查询 知识库 分析 敬请期待

## 所有样本

样本免疫组库基因使用情况，多样性等描述信息

配置信息 实验信息 表型信息 描述性统计

导出选中数据 显示 10 项结果 搜索:

点击每个字段右侧的箭头，可以令数据按照该字段进行升序或降序进行展示。

选择	样本	Reads	V alignment(%)	D alignment(%)	J alignment(%)	CDR3 种类数 (氨基酸)	CDR3 种类数 (核苷酸)	氨基酸 序列种类数	核苷酸 序列种类数
<input type="checkbox"/>	IS11815201710240673	808202	99.62	88.04	99.88	26221	21170	148222	114
<input type="checkbox"/>	IS11815201710240674	518250	99.6	86.14	99.79	24361	19871	133655	107
<input type="checkbox"/>	IS11815201710240675	18736	99.29	84.94	99.25	5818	5258	7912	719
<input type="checkbox"/>	IS11815201710240676	32033	99.34	86.32	99.84	3849	3139	8409	697
<input type="checkbox"/>	IS11815201710240677	67741	99.6	84.74	99.91	6812	5960	19014	161
<input type="checkbox"/>	IS11815201710240678	48950	99.62	87.05	99.71	4587	4043	12985	107
<input type="checkbox"/>	IS11815201710240679	324153	99.5	89.02	99.59	11522	9275	60826	476
<input type="checkbox"/>	IS11815201710240680	816582	99.43	82.45	99.82	21261	17057	191733	147
<input type="checkbox"/>	IS11815201710240681	892937	99.49	82.89	99.82	30499	24742	211118	166
<input type="checkbox"/>	IS11815201710240682	47579	99.31	84.31	99.69	5359	4581	14453	122

选择 样本 Reads V alignment(%) D alignment(%) J alignment(%) CDR3 种类数 (氨基酸) CDR3 种类数 (核苷酸) 氨基酸 序列种类数 核苷酸 序列种类数

## 导出选中数据



首页 项目 样本 检索查询 知识库 分析 敬请期待

### 所有样本

样本免疫组库基因使用情况，多样性等描述信息

配置信息 实验信息 表型信息 描述性统计

导出选中数据 (3) 显示 10 页 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 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

### 所有项目

**项目基本信息**

导出选中数据 显示 10 项结果 搜索:

选择	项目	个体数量	样本数量	受体类型	物种
<input type="checkbox"/>	HIV_vaccine	54	108	TCR&BCR	Rhesus_Monkey
<input type="checkbox"/>	Health	457	457	TCR	human
<input type="checkbox"/>	IgA_nephropathy	52	140	TCR&BCR	human
<input type="checkbox"/>	SLE	971	971	TCR	human
<input type="checkbox"/>	breast_cancer_research	16	79	TCR	human
<input type="checkbox"/>	colorectal_cancer	16	54	BCR	human

显示第 1 至 6 项结果, 共 6 项

## 4.4. 项目信息

项目的简单描述以及相关文献。

项目: breast\_cancer\_research

**配置信息**

个体

样本

**项目描述** breast\_cancer\_research  
 T lymphocytes infiltrate the microenvironment of breast cancer tumors and play a pivotal role in tumor immune surveillance. Relationships between the T-cell receptors (TCR) borne by T cells within tumors, in the surrounding tissues, and in draining lymph nodes are largely unexplored in human breast cancer. Consequently, information about the relative extent of possible T-cell exchange between these tissues is also lacking. Here, we have analyzed the TCR repertoire of T cells using multiplex PCR and high-throughput sequencing of the TCRβ chain in the tissues of tumor, adjacent nontumor, and axillary lymph nodes of breast cancer patients. T-cell repertoire diversity in tumors was lower than in lymph nodes, but higher than in nontumor tissue, with a preferential use of variable and joining genes. These data are consistent with the hypothesis that most of the T cells in tumors derive from the lymph node, followed by their expansion in tumor tissue. Positive nodes appeared to enhance T-cell infiltration into tumors and T-cell clonal expansion in lymph nodes. Additionally, the similarity in TCR repertoire between tumor and nontumor tissue was significantly higher in luminal-like, rather than basal-like, breast cancer. Our study elucidated the high heterogeneity of the TCR repertoire and provides potential for future improvements in immune-related diagnosis, therapy, and prognosis for breast cancer patients. Cancer Immunol Res; 5(2):148-56.2016 AACR

**序列类型** TCR  
**物种** human  
**相关文献** The different T-cell receptor repertoires in breast cancer tumors,draining lymph nodes and adjacent tissues PMID 28039116  
**发布日期** 2016-12-30



## 4.5. 个体信息

每个个体都有一个唯一对应的个体编号，编号以大写字母“II”开头。个体信息包括个体对应样本数、序列类型、物种等。

国家基因组 数据中心 项目总览 简体中文 登录/注册

---

**PIRD**

首页 项目 样本 检索查询 知识库 分析 敬请期待

项目 : breast\_cancer\_research

配置信息

**个体**

样本

### 项目中个体信息

导出选中数据 显示 10 项结果 搜索:

选择	个体	样本统计	序列类型	物种	性别
<input type="checkbox"/>	II955920160920BC0001A	5	TCR	human	Female
<input type="checkbox"/>	II955920160920BC0002A	6	TCR	human	Female
<input type="checkbox"/>	II955920160920BC0003A	5	TCR	human	Female
<input type="checkbox"/>	II955920160920BC0004A	5	TCR	human	Female
<input type="checkbox"/>	II955920160920BC0005A	4	TCR	human	Female

### 项目的性别组成

国家基因组 数据中心 项目总览 简体中文 登录/注册

---

**PIRD**

首页 项目 样本 检索查询 知识库 分析 敬请期待

项目 : breast\_cancer\_research

配置信息

**个体**

样本

### 项目中个体性别比例

性别分布

性别	数量	比例
女	25	100%

## 4.6. 样本信息

每个样本都有一个唯一对应的个体编号，编号以大写字母“IS”开头。

### 4.6.1. 样本配置信息

国家基因组 数据中心 项目总览 简体中文 登录/注册

---

 **PIRD**

---

首页 项目 样本 检索查询 知识库 分析 敬请期待

---

项目 : breast\_cancer\_research

配置信息

个体

**样本**

#### 样本配置信息

配置信息 实验信息 表型信息 描述性统计 

导出选中数据 显示 10 项结果 搜索:

选择	个体	样本	核酸类型(DNA/RNA)	物种	分子序列
<input type="checkbox"/>	II955920160920BC0006A	IS4041201609180001	DNA	human	TRB
<input type="checkbox"/>	II955920160920BC0011A	IS4041201609180002	DNA	human	TRB
<input type="checkbox"/>	II955920160920BC0013A	IS4041201609180003	DNA	human	TRB
<input type="checkbox"/>	II955920160920BC0002A	IS4041201609180004	DNA	human	TRB

## 4.6.2. 样本实验信息

国家基因组 数据中心 项目总览 简体中文 登录/注册

**PIRD**

首页 项目 样本 检索查询 知识库 分析 敬请期待

项目 : breast\_cancer\_research

配置信息  
个体  
**样本**

### 样本实验信息

配置信息 实验信息 表型信息 描述性统计

导出选中数据 显示 10 项结果 搜索:

选择	样本	个体	引物集	样本类型	分子序列	取样时间	测序平台
<input type="checkbox"/>	IS4041201609180001	II955920160920BC0006A	TRB-VJ-2015V1	PBMC	TRB	-	HiSeq
<input type="checkbox"/>	IS4041201609180002	II955920160920BC0011A	TRB-VJ-2015V1	PBMC	TRB	-	HiSeq
<input type="checkbox"/>	IS4041201609180003	II955920160920BC0013A	TRB-VJ-2015V1	PBMC	TRB	-	HiSeq
<input type="checkbox"/>	IS4041201609180004	II955920160920BC0002A	TRB-VJ-2015V1	PBMC	TRB	-	HiSeq

## 4.6.3. 样本表型信息

国家基因组 数据中心 项目总览 简体中文 登录/注册

**PIRD**

首页 项目 样本 检索查询 知识库 分析 敬请期待

项目 : breast\_cancer\_research

配置信息  
个体  
**样本**

### 样本对应的个体的表型信息

配置信息 实验信息 表型信息 描述性统计

导出选中数据 显示 10 项结果 搜索:

选择	样本	性别	年龄	民族	身高 (厘米)	体重 (千克)	体质指数(BMI)	血压	空腹血糖	个体状态	取样日期
<input type="checkbox"/>	IS4041201609180001	Female	52	han	-	-	-	-	-	breast cancer	-
<input type="checkbox"/>	IS4041201609180002	Female	47	han	-	-	-	-	-	breast cancer	-
<input type="checkbox"/>	IS4041201609180003	Female	47	han	-	-	-	-	-	breast cancer	-

#### 4.6.4. 样本描述性统计信息

描述性统计信息反映了样品的免疫组库测序的一些统计数据，包括 Reads 数量、基因对比率、CDR3 序列的种类、CDR3 的多样性（香农熵）等信息。

国家基因库 数据中心 项目总览 简体中文 登录/注册

---

**PIRD**

---

首页 项目 样本 检索查询 知识库 分析 敬请期待

项目 : breast\_cancer\_research

配置信息 个体 样本

### 样本免疫组库基因使用情况，多样性等描述信息

配置信息 实验信息 表型信息 描述性统计

导出选中数据 显示 10 项结果 搜索:

选择	样本	Reads	V alignment(%)	D alignment(%)	J alignment(%)	CDR3 种类数 (氨基酸)	CDR3 种类数 (核苷酸)	氨基酸序列种类数
<input type="checkbox"/>	IS4041201609180001	11173629	70.87	34.05	70.72	166229	260024	1045810
<input type="checkbox"/>	IS4041201609180002	3462937	80.31	49.86	80.2	99514	151926	517437
<input type="checkbox"/>	IS4041201609180003	21888651	76.12	46.74	76.07	373435	635263	2550761

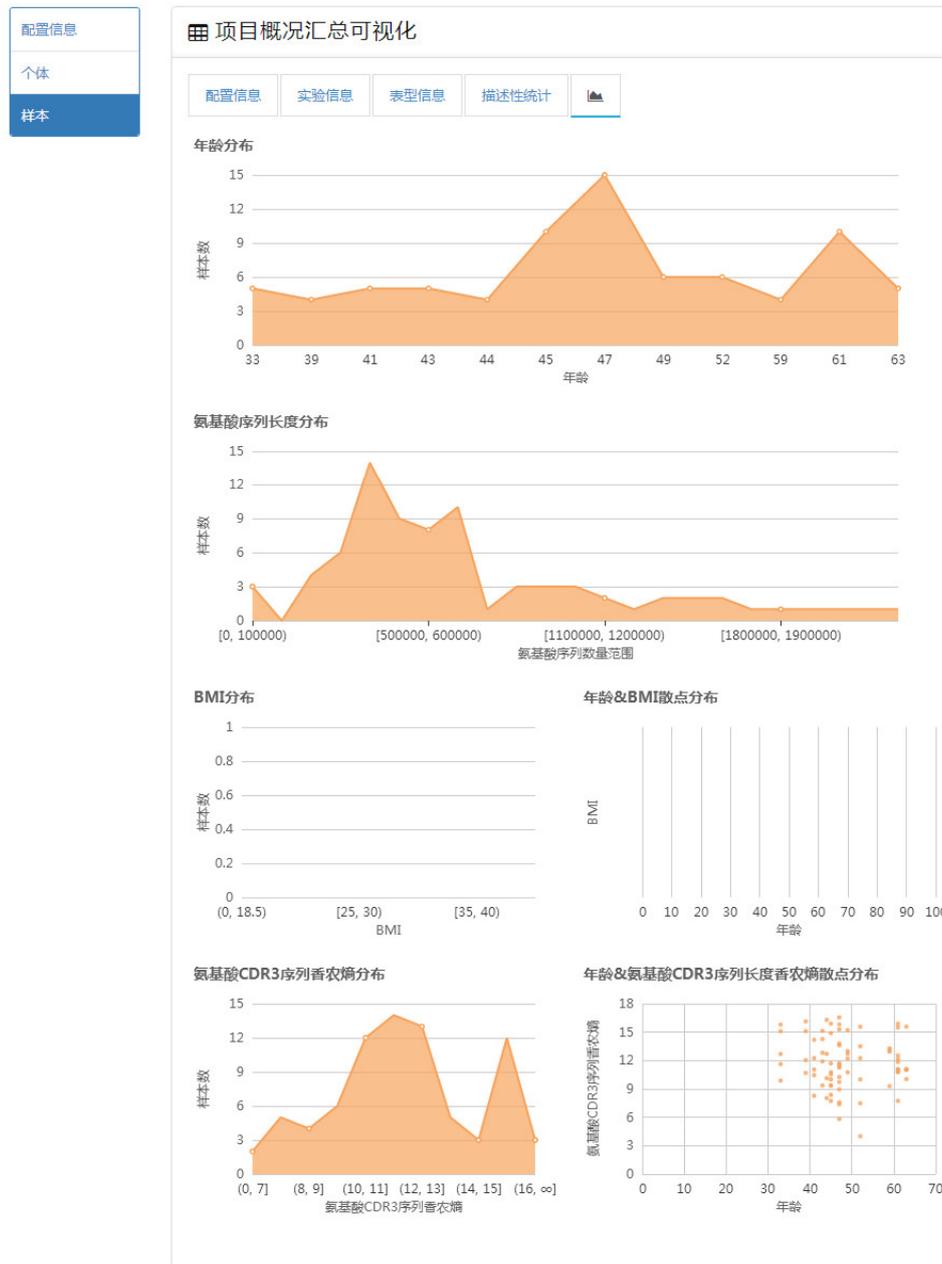
## 4.6.5. 项目全部样本概况汇总展示

国家基因组 数据中心 项目总览 简体中文 登录/注册

**PIRD**

首页 项目 样本 检索查询 知识库 分析 敬请期待

项目：breast\_cancer\_research



## 4.6.6. 样本详情

### 样本的序列信息

包括克隆的具体序列，克隆频率，基因比对情况等。

国家基因库 数据中心 项目总览 简体中文 wanglonglong

# PIRD

首页 项目 样本 检索查询 知识库 分析 敬请期待

项目 : colorectal\_cancer/个体 : I11815201711110001/样本 : IS11815201710240673

序列详情

- V 使用数量
- J 使用数量
- V-J 使用数量
- 前10条克隆
- CDR3 长度

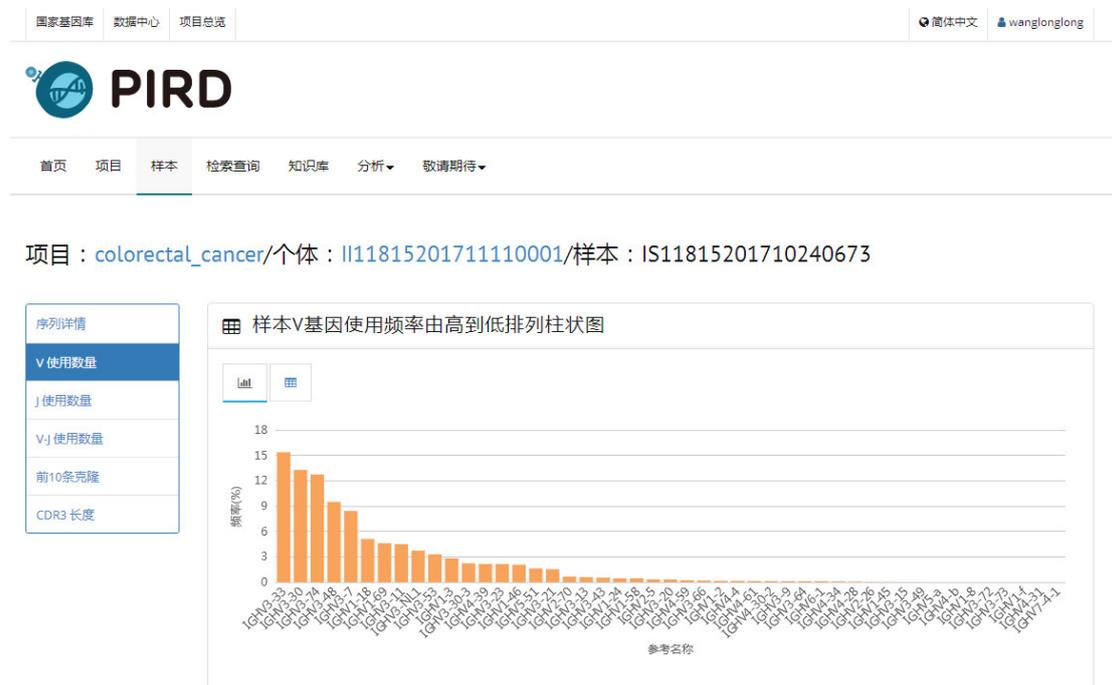
#### 样本中全部序列的具体信息

导出当前表格 显示 10 项结果

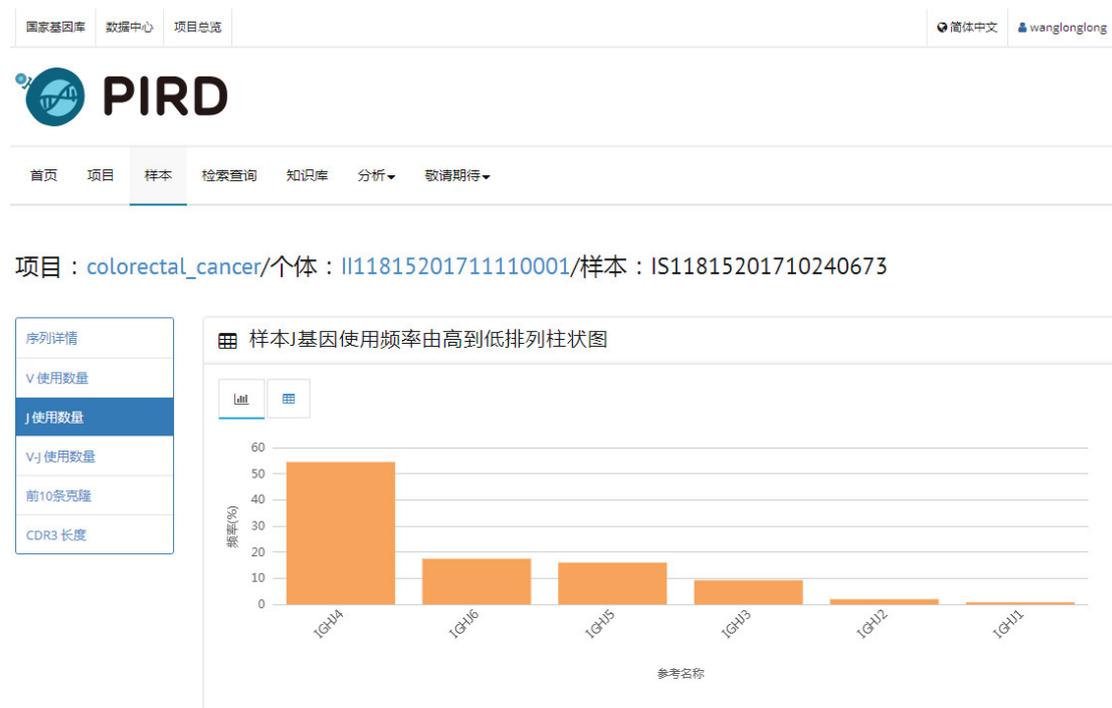
序列类型	频率(%)	V 名称	D 名称	J 名称	CDR3 起始位置	CDR3 终止位置	CDR3 核苷酸序列
in-frame	0.000117774948561791	IGHV1-24*01	IGHD3-22*01	IGHJ4*02	88	144	TGTGCGAGAGATCTCGACTATGAGAGTTATC
in-frame	0.000117774948561791	IGHV1-24*01	-	IGHJ4*02	88	126	TGTGCAACAGATCTCACGGACGATGAACTTC
in-frame	0.000117774948561791	IGHV1-24*01	-	IGHJ4*02	88	126	TGTGCAACAGATCTCACGGACGATGAACTTC

## VJ 基因的使用情况

柱状图由高到低排序展示各 V 基因的使用频率。



柱状图由高到低排序展示各 J 基因的使用频率。



热图展示 V-J 基因组合的使用频率

热图的右侧颜色示例上下界可以通过上下拖拽三角形箭头的位置进行调整，用于选择展示的 VJ 基因组合的频率范围。

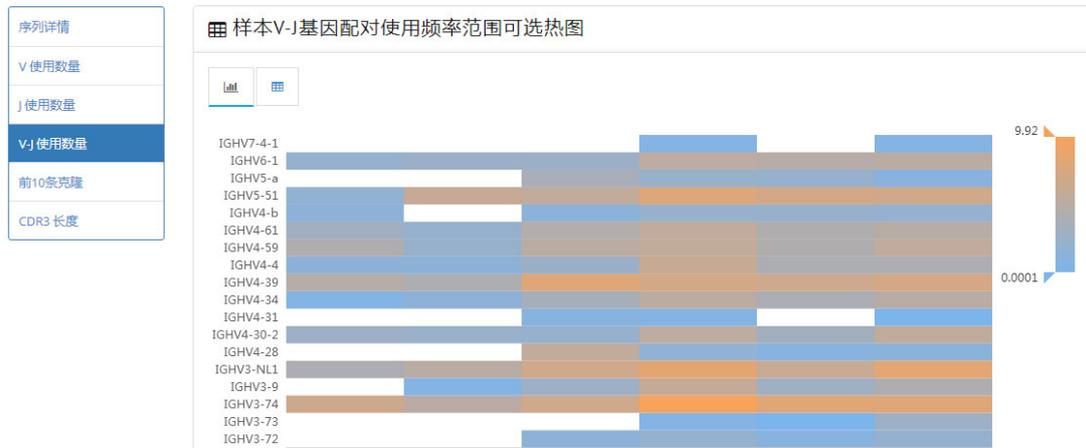
国家基因库 数据中心 项目总览 简体中文 wanglonglong

---

**PIRD**

首页 项目 样本 检索查询 知识库 分析 敬请期待

项目 : colorectal\_cancer/个体 : II11815201711110001/样本 : IS11815201710240673



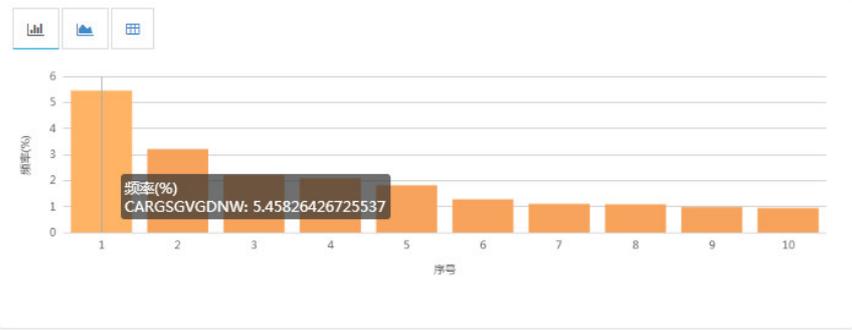
频率前 10 克隆

频率前十克隆的频率分布和排行

项目 : colorectal\_cancer/个体 : I11815201711110001/样本 : IS11815201710240673

- 序列详情
- V 使用数量
- J 使用数量
- V-J 使用数量
- 频率前 10 克隆**
- CDR3 长度

样本频率前10的克隆由高到低排列柱状图

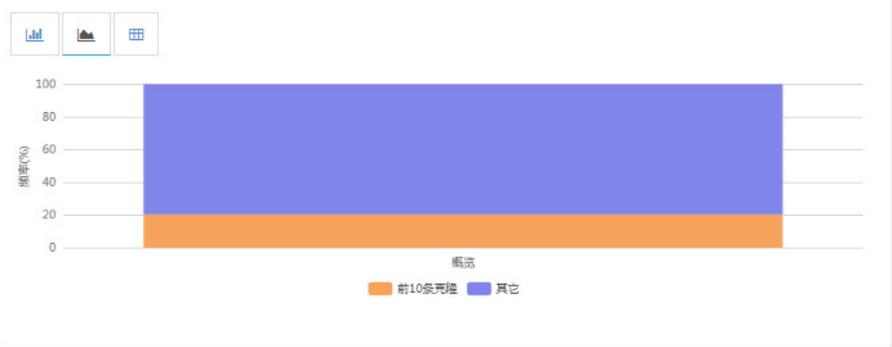


频率前十克隆占有所有克隆的百分比

项目 : colorectal\_cancer/个体 : I11815201711110001/样本 : IS11815201710240673

- 序列详情
- V 使用数量
- J 使用数量
- V-J 使用数量
- 频率前 10 克隆**
- CDR3 长度

样本频率前10的克隆频率总和占比堆积图



## CDR3 长度

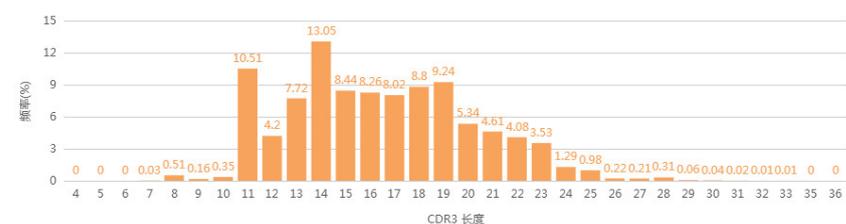


首页 项目 样本 检索查询 知识库 分析 敬请期待

项目 : colorectal\_cancer/个体 : I11815201711110001/样本 : IS11815201710240673

序列详情
V 使用数量
J 使用数量
VJ 使用数量
频率前 10 克隆
<b>CDR3 长度</b>

样本CDR3长度 (氨基酸残基数) 分布柱状图



## 五、检索查询

### 5.1. 检索项目



首页 项目 样本 检索查询 知识库 分析 敬请期待

#### 检索查询

<b>检索项目</b>
检索样本
检索克隆
BLAST

项目 :

样本数量 最小值 ~ 最大值

最小值和最大值必须是非负整数。

物种 -- 请选择 --

受体类型 -- 请选择 --

提交

## 5.2. 检索样本



PIRD

首页 项目 样本 检索查询 知识库 分析 敬请期待

### 检索查询

- 检索项目
- 检索样本
- 检索克隆
- BLAST

取样时间

年龄  ~   
① 最小值和最大值必须是非负整数。

性别

分子序列

测序平台

序列类型

体质指数(BMI)  ~   
① 最小值和最大值必须是非负的浮点数。

个体状态

CDR3香农熵(氨基酸)  ~   
① 最小值和最大值必须是非负的浮点数。

## 5.3. 检索克隆

目前版本，克隆只能通过核苷酸序列进行检索，如果需要检索氨基酸序列，可以通过 BLAST 模块选择 BLASTP 进行检索。

### 检索查询

- 检索项目
- 检索样本
- 检索克隆
- BLAST

<b>取样时间</b>	<input type="text"/>
<b>年龄</b>	<input type="text" value="最小值"/> ~ <input type="text" value="最大值"/> <small>① 最小值和最大值必须是非负整数。</small>
<b>性别</b>	<input type="text" value="-- 请选择 --"/>
<b>分子序列</b>	<input type="text" value="-- 请选择 --"/>
<b>测序平台</b>	<input type="text" value="-- 请选择 --"/>
<b>序列类型</b>	<input type="text" value="-- 请选择 --"/>
<b>体质指数(BMI)</b>	<input type="text" value="最小值"/> ~ <input type="text" value="最大值"/> <small>① 最小值和最大值必须是非负的浮点数。</small>
<b>个体状态</b>	<input type="text" value="-- 请选择 --"/>
<b>CDR3香农熵(氨基酸)</b>	<input type="text" value="最小值"/> ~ <input type="text" value="最大值"/> <small>① 最小值和最大值必须是非负的浮点数。</small>
<b>序列选择(全长/CDR3)</b>	<input type="text" value="-- 请选择 --"/> <small>① 必填项</small>
<b>核苷酸序列</b>	<input type="text"/> <small>① 必填项</small>

## 5.4. BLAST 检索说明

非完整序列可以通过 BLAST 进行检索，点击 BLAST 标签，界面会跳转到基因库统一的 BLAST 检索平台。

## 检索查询

- 检索项目
- 检索样本
- 检索克隆
- BLAST**

取样时间 年龄  ~ **!** 最小值和最大值必须是非负整数。性别 分子序列 测序平台 序列类型 体质指数(BMI)  ~ **!** 最小值和最大值必须是非负的浮点数。个体状态 CDR3香农熵(氨基酸)  ~ **!** 最小值和最大值必须是非负的浮点数。序列选择(全长/CDR3) **!** 必填项核苷酸序列 **!** 必填项

## BLAST: 国家基因库序列搜索服务

国家基因库 (CNGB) 序列搜索服务整合了国家基因库所有项目的可用数据, 致力于提供广泛涵盖各个研究领域的、高效便捷的序列搜索服务。

<b>BLASTN</b> 核苷酸 → 核苷酸	<b>BLASTX</b> 核苷酸 (自动翻译) → 蛋白质	<b>BLASTP</b> 蛋白质 → 蛋白质
	<b>TBLASTN</b> 蛋白质 → 核苷酸 (自动翻译)	

### BLAST新闻

- 2017-11-30**  
千种植物项目 (OneKP) 的所有 (1328个) 单样本的核苷酸序列和蛋白质序列数据现已全部整合到BLAST服务。
- 2017-11-02**  
任务结果页面上, 新增下载比对中的序列的功能。
- 2017-06-29**  
序列搜索服务新增NCBI BLAST数据库、千种植物数据库 (OneKP)、免疫数据库 (PIRD) 和千种鱼转录组项目 (FishT1K) 4个CNGB项目数据。BLAST服务公开版本发布。
- 2017-06-15**  
序列搜索服务完成并通过所有功能组件测试。测试版本开放内部测试。
- 2017-05-26**  
序列搜索服务后台计算平台完成设计。
- 2017-05-03**  
序列搜索服务开始开发。

### 生物多样性

-  10KP
-  1KITE
-  B10k
-  BCGD
-  Fish T1K
-  Millet
-  MLGD
-  MT10K
-  1KP

### 健康与疾病

-  BDDB
-  CCDB
-  DHGV
-  DISSECT
-  GDRD
-  GeMap
-  HIFTD
-  HMD
-  ICGC
-  PIRD
-  PVD
-  SCDB

### 模块服务

-  Biomigo
-  BLAST
-  CNSA
-  GIGA<sup>TM</sup>

## 检索示例

以氨基酸序列的检索为例, 我们先从 SLE 项目中找到一条频率比较高的 CDR3 克隆的氨基酸序列: CASSDWGNTEAFF, 下一步通过 BLASTP 查看。

**第一步：**在查询序列之前要先选择用于进行比对的数据库，界面右侧的操作框内可以选择数据库的搜索范围，这里选择“PIRD：免疫数据库”。

The image shows the BLAST web interface. On the left, the 'BLAST选项' (BLAST Options) panel is visible, with the 'BLASTP' tab selected. The '输入选项' (Input Options) section includes a text area for the query sequence and a '选择文件' (Select File) button. The '比对数据库' (Search Database) section has a dropdown menu. The '通用选项' (General Options) section includes settings for '程序类型' (Program Type) set to BLASTP, '最大目标序列数' (Maximum Number of Targets) set to 100, '期望值' (Expectation Value) set to 10, '字长' (Word Size) set to 6, '空位罚分' (Gap Penalties) set to '新建: 11, 拓展: 1', and '打分矩阵' (Scoring Matrix) set to BLOSUM62. On the right, the '搜索/检索数据库' (Search/Select Database) panel is highlighted with a red box. It features a '快速搜索' (Quick Search) section with a dropdown menu currently showing 'PIRD: 免疫数据库' selected. Below this is a search input field with a magnifying glass icon and the text '点击搜索' (Click Search) with a red arrow pointing to the icon. At the bottom, the '数据库搜索结果' (Database Search Results) section shows a list of databases: 'PIRD Knowledge Repository PIRD db', 'PIRD PIRD db', and 'OneKP ONEKP db'.

## 第二步：选定比对数据库搜索结果。



首页 BLAST ▾ 任务 帮助

### BLAST

BLASTN BLASTP BLASTX TBLASTN TBLASTX

#### BLAST选项

新窗口打开? 提交

**输入选项**

查询序列文本 或 或者查询序列文件

选择文件 未选择任何文件

任务描述

比对数据库 **PIRD [PIRD][db]** 选择好的比对数据库在此显示  
从右边面板搜索/检索并选择比对数据库。点击该单元格中的标签删除选中的数据库。

查询序列的子区域

**通用选项**

程序类型 BLASTP

最大目标序列数 100

期望值 10

字长 6

空位罚分 新建: 11, 拓展: 1

打分矩阵 BLOSUM62

基于组成成分的统计方法 Conditional composition-based score adjustment

更多选项 ▾

#### 搜索/检索数据库

快速搜索 ▾

项目 PIRD: 免俗数据库

分类 -- 请选择 --

级别

关键字

高级检索 (敬请期待)

#### 数据库搜索结果

PIRD Knowledge Repository **PIRD db**

PIRD **PIRD db**

OneKP **ONEKP db**

1. 点击搜索结果中找到的比对数据库PIRD

第三步：输入待查询的序列，选择比对参数，提交任务。

**BLAST**

首页 BLAST 任务 帮助

BLAST

BLASTN **BLASTP** BLASTX TBLASTN TBLASTX

3.提交任务

BLAST选项  新窗口打开? **提交**

输入选项

查询序列文本 或者查询序列文件 **CASSDWGNTAEFF**

1.输入待查询序列

选择文件 未选择任何文件

任务描述

比对数据库 **P9559201609201 [PIRD][project]**

从右边面板搜索/检索并选择比对数据库。点击该单元格中的标签删除选中的数据库。

查询序列的子区域

通用选项

2.选择比对参数，也可默认

程序类型 **BLASTP**

最大目标序列数 **100**

期望值 **10**

字长 **6**

空位罚分 **新建: 11, 拓展: 1**

打分矩阵 **BLOSUM62**

基于组成成分的统计方法 **Conditional composition-based score adjustment**

更多选项

搜索/检索数据库

快速搜索

项目 **PIRD: 免疫数据库**

分类级别 -- 清洗 --

关键字

高级检索 (敬请期待)

数据库搜索结果

PIRD Knowledge Repository **PIRD db**

PIRD **PIRD db**

P9559201608102 **PIRD project**

I19559201608105533 **PIRD individual**

I19559201608105534 **PIRD individual**

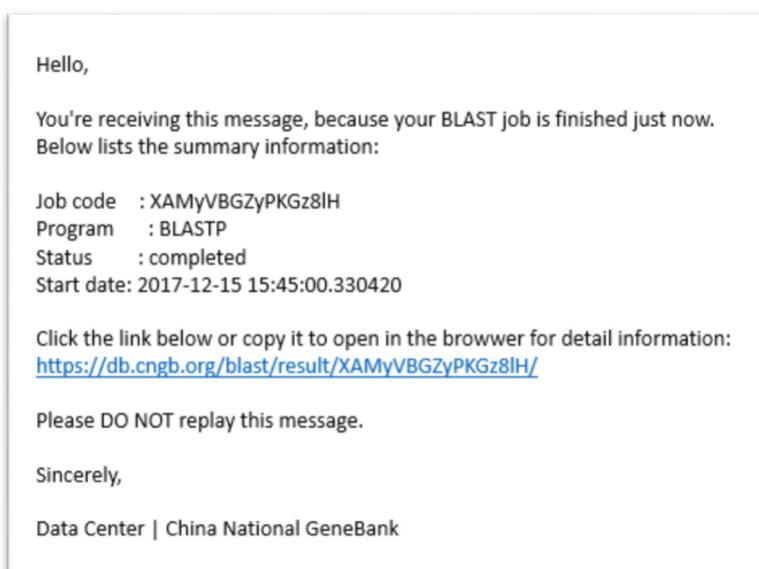
I19559201608105535 **PIRD individual**

I19559201608105539 **PIRD individual**

I19559201608105548 **PIRD individual**

I19559201608105551 **PIRD individual**

第四步：等待 BLASTP 比对结果，比对任务完成后会收到通知邮件。



第五步：查看比对结果，在详细信息中可以看到匹配度较高的序列所在项目的样品编号。

**BLAST**

首页 BLAST 任务 帮助

任务结果: XAMyVBGZyPKGz8IH

数据库搜索结果	
任务描述	- (失效日期为2017-12-18 15:45:06.846321)
状态	完成
任务结果	<a href="#">Pairwise</a> <a href="#">BLAST XML</a> <a href="#">Tabular</a> <a href="#">Comma-separated values</a> <a href="#">BLAST archive (ASN.1)</a> <a href="#">Seqalign (JSON)</a> <a href="#">Multiple-file BLAST JSON</a> <a href="#">Multiple-file BLAST XML2</a> <a href="#">Single-file BLAST JSON</a> <a href="#">Single-file BLAST XML2</a>
当前查询序列	Query_1: unnamed protein product

摘要	
查询序列ID	Query_1
数据库名称	pirid_P9559201609201
插叙序列描述	unnamed protein product
数据库描述	P9559201609201[pirid][project]
查询序列长度	13
程序	BLASTP 2.6.0+

任务选项配置	
程序类型	blastp
打分矩阵	BLOSUM62
期望值	10
基于组成成分的统计方法	2
SEG过滤器	no
字长	6
新建空位罚分	11
拓展空位罚分	1
最大目标序列数	100

比对序列描述列表				下载选中	下载所有
#	比对序列描述	打分	期望值	匹配点	
<input type="checkbox"/>	gnl pird IS4041201609180055 2b6f8aec29a67470	29.6	0.021	92%	
<input type="checkbox"/>	gnl pird IS4041201609180017 2b6f8aec29a67470	29.6	0.021	92%	
<input type="checkbox"/>	gnl pird IS4041201609180066 444b5da186dee03c	28.9	0.035	92%	
<input type="checkbox"/>	gnl pird IS4041201609180041 444b5da186dee03c	28.9	0.035	92%	
<input type="checkbox"/>	gnl pird IS4041201609180040 444b5da186dee03c	28.9	0.035	92%	
<input type="checkbox"/>	gnl pird IS4041201609180031 444b5da186dee03c	28.9	0.035	92%	
<input type="checkbox"/>	gnl pird IS4041201609180025 444b5da186dee03c	28.9	0.035	92%	
<input type="checkbox"/>	gnl pird IS4041201609180008 444b5da186dee03c	28.9	0.035	92%	
<input type="checkbox"/>	gnl pird IS4041201609180059 e548cbccf7a14b96	28.9	0.038	92%	
<input type="checkbox"/>	gnl pird IS4041201609180010 e548cbccf7a14b96	28.9	0.038	92%	
<input type="checkbox"/>	gnl pird IS4041201609180045 4f30c2d1834de5ad	28.9	0.041	92%	
<input type="checkbox"/>	gnl pird IS4041201609180067 a084002c1ae4faa8	28.9	0.042	92%	
<input type="checkbox"/>	gnl pird IS4041201609180021 4f0c976884415bbb	28.5	0.052	92%	
<input type="checkbox"/>	gnl pird IS4041201609180065 d36396f8d1fc77ec	28.5	0.054	92%	
<input type="checkbox"/>	gnl pird IS4041201609180021 20fafda33e94e8cf	28.5	0.054	92%	
<input type="checkbox"/>	gnl pird IS4041201609180017 20fafda33e94e8cf	28.5	0.054	92%	
<input type="checkbox"/>	gnl pird IS4041201609180018 01db8b1b50e2585f	28.5	0.064	92%	
<input type="checkbox"/>	gnl pird IS4041201609180064 018075805b31c5ca	28.1	0.069	92%	
<input type="checkbox"/>	gnl pird IS4041201609180053 018075805b31c5ca	28.1	0.069	92%	
<input type="checkbox"/>	gnl pird IS4041201609180022 018075805b31c5ca	28.1	0.069	92%	
<input type="checkbox"/>	gnl pird IS4041201609180006 018075805b31c5ca	28.1	0.069	92%	
<input type="checkbox"/>	gnl pird IS4041201609180041 3e901dd715645900	28.1	0.071	92%	
<input type="checkbox"/>	gnl pird IS4041201609180034 3e901dd715645900	28.1	0.071	92%	

比对结果对应的样品编号

Browser
Comming soon.

Alignments					
gnl pird IS4041201711110137-8a4c11f303f0b238 CDR3; DNA; TCR; human; 0.00% Length: 13 Matches: 1 <a href="#">Back to Descriptions</a>					
Range 1: 1 - 13					
<b>Score</b>	<b>Except</b>	<b>Identities</b>	<b>Positives</b>	<b>Gaps</b>	
30.4 bits(67)	0.08	13/13(100%)	13/13(100%)	0/13(0%)	
Query	1	CASSDWGNTAEFF	13		
		CASSDWGNTAEFF			
Sbjct	1	CASSDWGNTAEFF	13		
gnl pird IS4041201711110040-8a4c11f303f0b238 CDR3; DNA; TCR; human; 0.00% Length: 13 Matches: 1 <a href="#">Back to Descriptions</a>					
Range 1: 1 - 13					
<b>Score</b>	<b>Except</b>	<b>Identities</b>	<b>Positives</b>	<b>Gaps</b>	
30.4 bits(67)	0.08	13/13(100%)	13/13(100%)	0/13(0%)	
Query	1	CASSDWGNTAEFF	13		
		CASSDWGNTAEFF			
Sbjct	1	CASSDWGNTAEFF	13		

## 六、在线分析

用户在线的数据分析模块可以对项目数据以及样本数据自主进行可视化分析,保存及导出精致的图片。可视化的过程需要依次执行三个步骤：数据源-数据表-可视化。

### 6.1. 数据源

首先需要选择数据的来源,每个用户拥有 100Mb 的数据存储空间,数据的来源可以是用户自己上传的数据,也可以从数据库的公共数据中选择的数据源。

#### 6.1.1. 用户自己上传的数据。

数据上传过程会有详细的指引界面,帮助用户按照正确的格式进行数据上传。



目前上传功能只接受EXCLE.CSV (以逗号[,]分隔).TXT (以制表符Tab分隔)三种文件格式的数据上传,上传的数据需严格按照上传数据模板的格式,具体数据格式见模板(可下载)。

sample表格示例:

Sample	Sequence type	Species	Locus	Primer set	Type	Sampling date	Platform	Gender	Age	Race	Height(cm)	Weight(cm)
IS4041201608040001	DNA	human	TCR	TRB-VJ-2015V1	PBMC	-	Hiseq	Female	24	han		

cdr3\_statistics表格示例:

V usage  J usage  VJ usage  Top10 clone  CDR3 length

Sample	Reference name	Usage frequency	Usage number
IS4041201608040001	TRBV16	0.00485745	29

[sample模板下载](#) [cdr3\\_statistics模板下载](#)

下一步

个人上传数据也需要以项目为单位进行,首先需要填入项目的相关信息。

1 协议      2 采集信息      3 上传文件      4 工作表设置

\* 已建项目:

项目名称:

描述:

样本类型:

\* 物种:

相关文章:

PMID:

发布日期:

1 协议      2 采集信息      3 上传文件      4 工作表设置

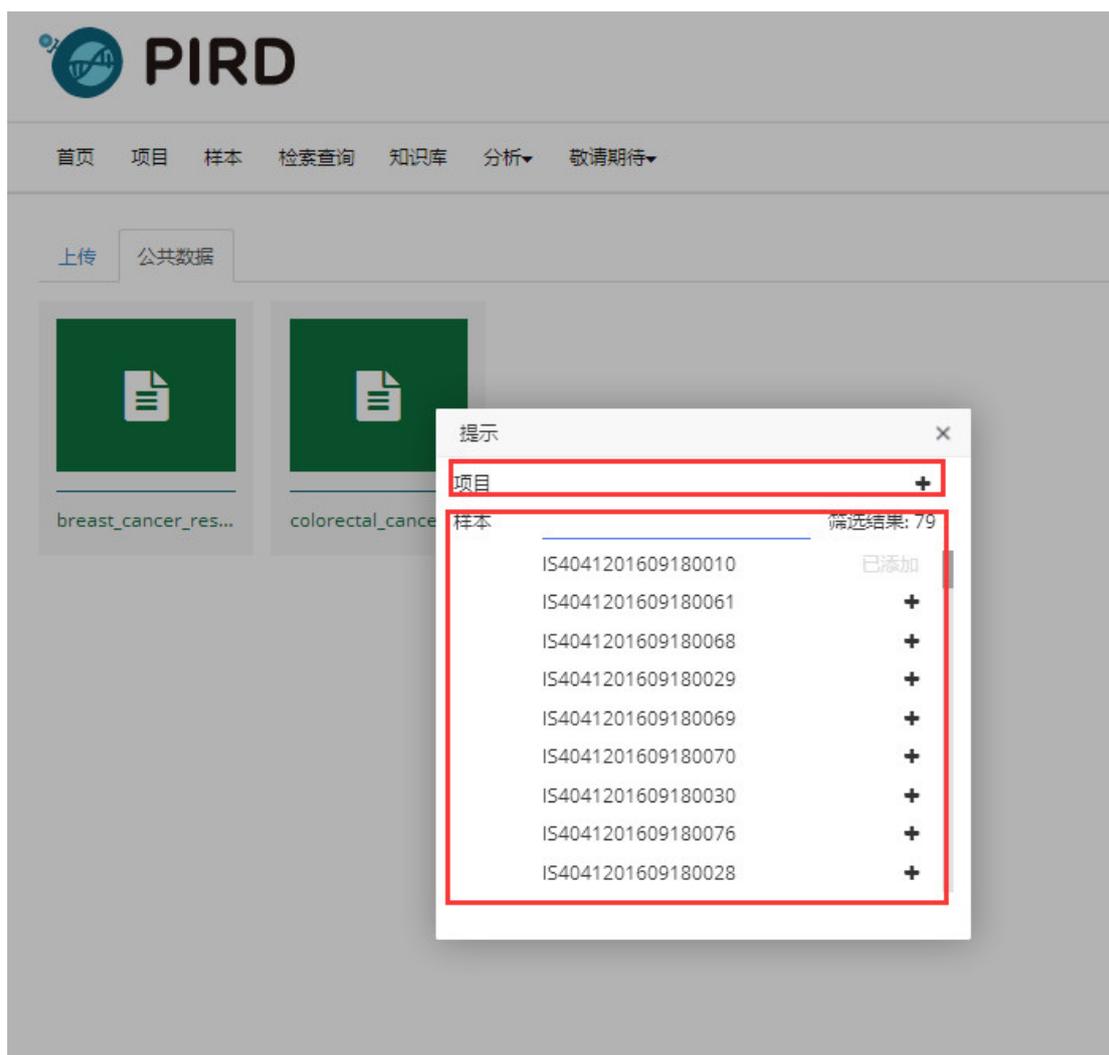
[点击上传文件](#)  
 支持Excel,Txt和CSV文件 (单个文件最大2M)  
 暂不支持文件批量上传,默认识别第一个sheet文件

1.请上传有标准行列的一维数据表格。(有合并单元格的数据请处理后再上传,否则可能出现表头识别有误)

2.日期字段需包含年月日(如2016/1/1),或年月日时分秒。(如2016/1/1 00:00:00)

### 6.1.2. 数据库现有的公共数据。

公共数据可以选择项目 (project) 和样本 (sample)



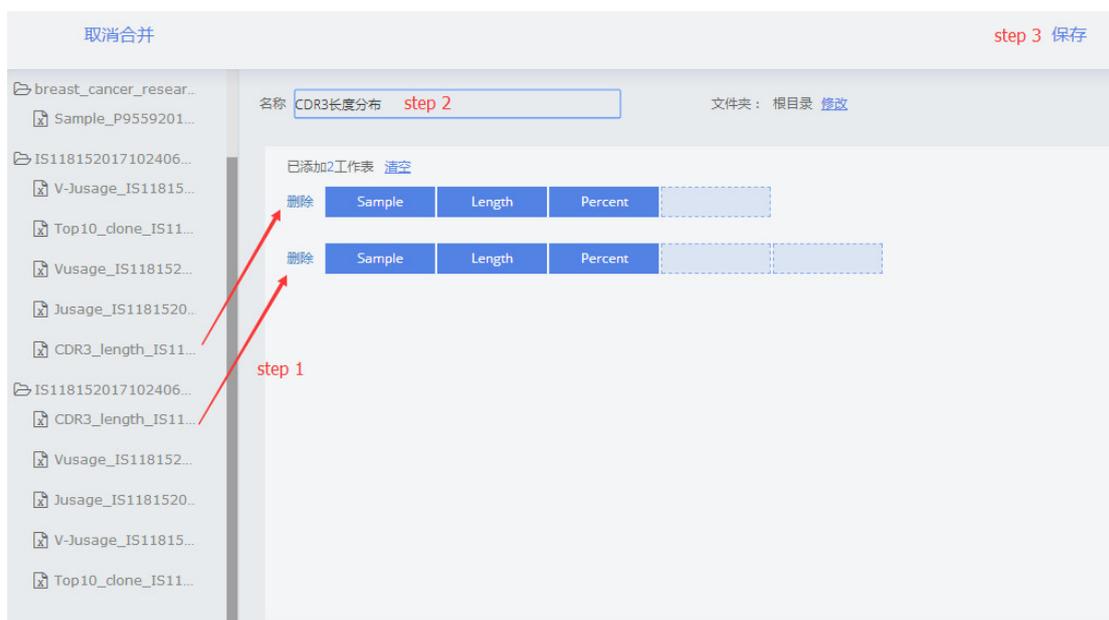
项目数据源里面包含了一个项目中所有样本的统计数据, 可用于群体及群体间的数据展示和比较。

样本数据源包含某个样本 VJ 基因使用, CDR3 长度等多个方面的信息, 可以就某个方面比较不同样本的差别。

## 6.2. 数据表

选定数据源后, 需要进一步生成用于作图的数据表, 数据表有追加合并的功能, 可以将多个数据源依次追加到同一个数据表中用于后续的作图。从左侧数据源栏侧拖拽用于合并的数据

到右侧即可追加数据并保存成为数据表。



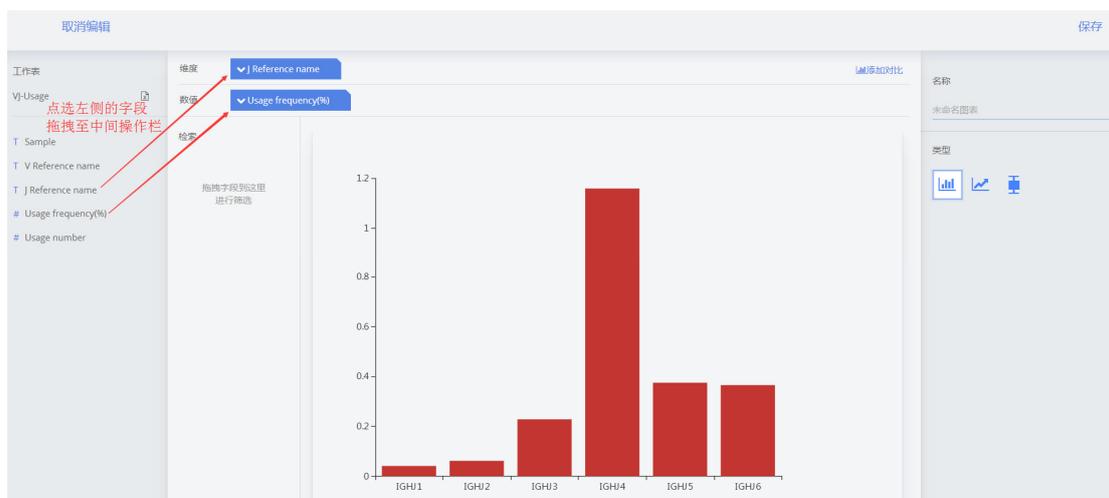
## 6.3. 可视化

界面左侧是数据表的所有字段，中间的操作栏中有**维度（横坐标）、对比、数值（纵坐标）**和**筛选**四个栏目，右侧可以对作图命名及选择展示图标的类型（柱状图，折线图，箱线图）。



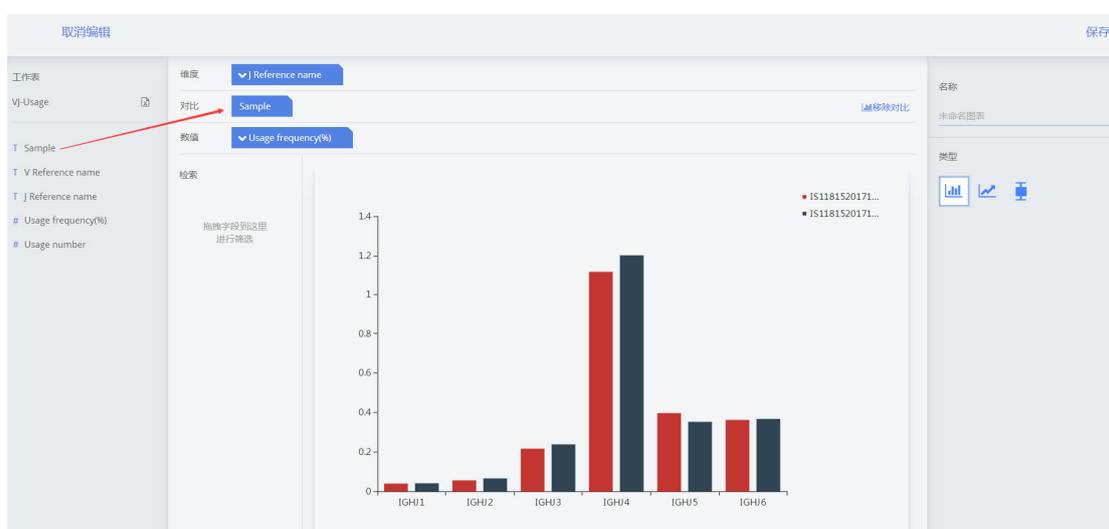
绘制简单的柱状图

最基本的绘图至少需要填入**维度（横坐标）**和**数值（纵坐标）**两栏，操作方式为点选左侧的字段拖拽至中间对应的操作栏。



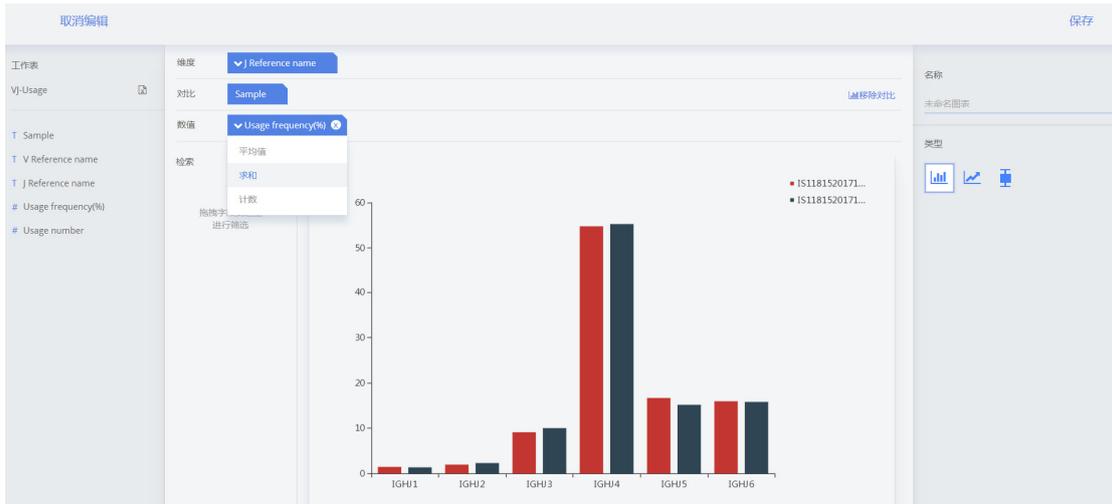
### 添加对比:

首先点选中间操作栏的“添加对比”按钮，然后拖拽需要进行对比的字段至对比操作栏。



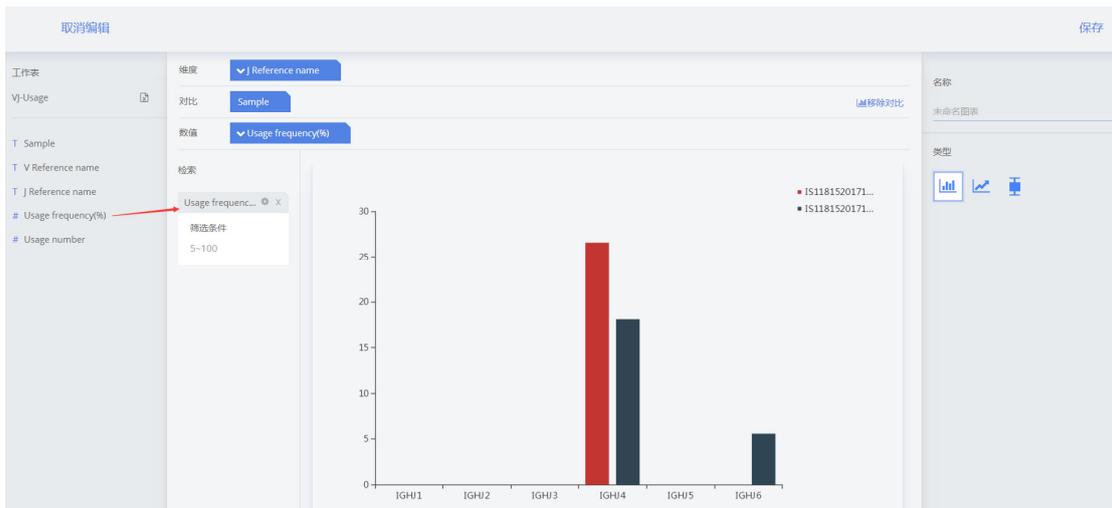
### 数值类型：

数值展示可以选择平均值，求和以及计数。



## 数据筛选：

点选拖拽要筛选显示的字段到中间筛选操作栏。



## 折线图

与柱状图类似。



## 箱线图

箱线图一般用于展示**多个样本组**比较的数据, 一般利用项目生成的数据表比较时会使用此种作图类型。例如下图展示了来自不同组织 (Type) 和不同疾病状态 (Status) 样本的香农熵, 可以看出“肿瘤 (tumor)”和“良性肿瘤 (benign tumor)”的 CDR3 香农熵的分布存在差异。这里同样可以加入多个字段筛选需要展示的数据。



## 七、知识库

知识库目前为内部华大免疫实验室内部使用，暂不对外开放。

PIRD

首页 项目 样本 检索查询 知识库 分析 敬请期待

### 知识库

参考序列

显示 10 项结果

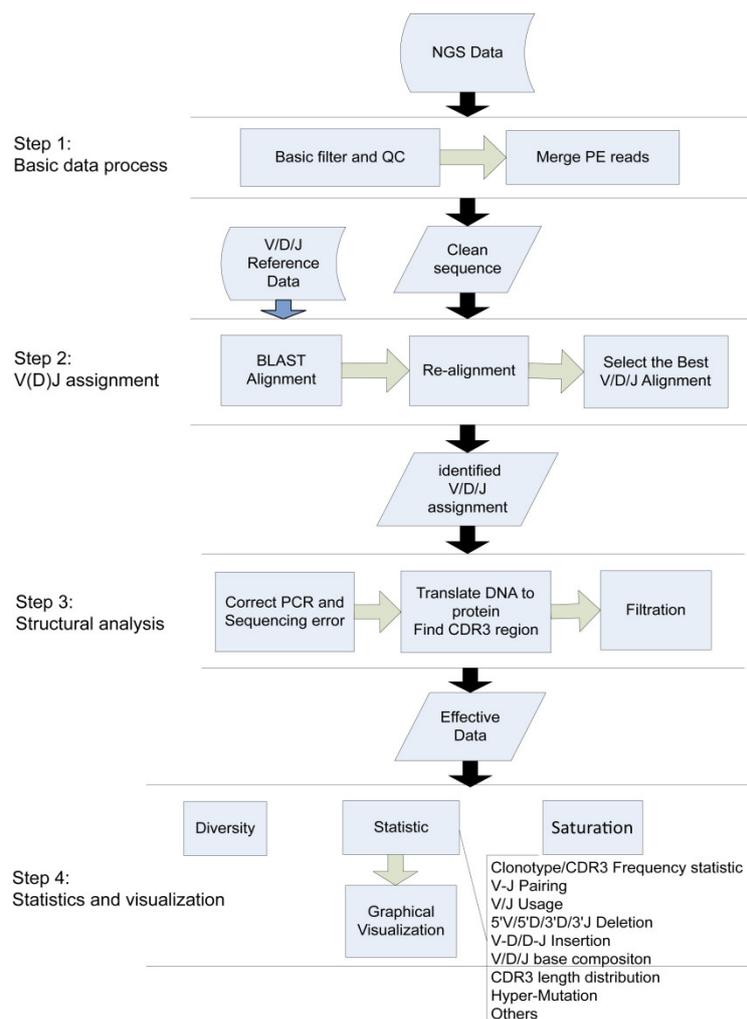
ICD名称	疾病名称	抗原	HLA	链式	长度	CDR3	V基因	D基因	J基因	测序平台	物种	起源
-------	------	----	-----	----	----	------	-----	-----	-----	------	----	----

Help Please refer to the FAQ on the Help page(<https://db.cngb.org/pird/help/>) for more assistance.

## 附录一、数据获取方法

### Method

The main analysis tool used for the raw repertoire sequencing data is IMonitor. IMonitor utilizes realignment to identify accurately V(D)J genes and alleles, and identify the CDR3 region, deletion/insertion of VDJ and then translate nucleotide into amino acid. It provides general



adaptation for sequences from all receptor chains of different species and outputs useful statistics and visualizations. Besides, PCR and sequencing errors are corrected for all sequences. [Learn more: https://github.com/zhangwei2015/IMonitor](https://github.com/zhangwei2015/IMonitor)