

## Supporting Information

### 1 Development and Testing of Druglike Screening Libraries

2 Junmei Wang\*, Yubin Ge, Xiang-Qun Xie

3 Department of Pharmaceutical Sciences, The University of Pittsburgh, 3501 Terrace St

4 Pittsburgh, PA 15261

5 \*Corresponding author: Junmei Wang

6 E-mail: juw79@pitt.edu Phone: +1-412-383-3268

7

8 **Table S1.** List of hit numbers for all drugs.

| SS | All     |         | GPCR    |         | Ion Channel |         | Protease |         | Kinase |         |
|----|---------|---------|---------|---------|-------------|---------|----------|---------|--------|---------|
|    | Ave     | Max     | Ave     | Max     | Ave         | Max     | Ave      | Max     | Ave    | Max     |
| 85 | 58.4    | 412.8   | 63.7    | 319.5   | 98.7        | 444.8   | 64       | 428.3   | 15.1   | 165.7   |
| 80 | 113.8   | 747.8   | 113.4   | 435.8   | 193.7       | 677.9   | 138.3    | 751.5   | 28.6   | 220.4   |
| 75 | 231.6   | 1310.6  | 259.2   | 757.3   | 413.5       | 1195.2  | 320.4    | 1517.9  | 60.9   | 340.8   |
| 70 | 524.9   | 2436.8  | 675.8   | 1759.2  | 1036.4      | 2870.6  | 801.6    | 3175    | 143.4  | 687     |
| 65 | 1248    | 4707    | 1840.2  | 4488.8  | 2588.3      | 6852    | 2031.9   | 6554.8  | 348.8  | 1463.2  |
| 60 | 2979.4  | 9536.9  | 4832    | 10444.6 | 6077.3      | 14479.8 | 5111.2   | 14019.5 | 876.8  | 3275.8  |
| 55 | 7438.2  | 20629   | 12830.8 | 24604.5 | 14510.3     | 31072.4 | 13104.9  | 30663.3 | 2431.7 | 7828    |
| 50 | 17799.2 | 43719.6 | 31959.1 | 54183.9 | 32540.7     | 62524   | 31705    | 64830.6 | 6746.1 | 18322.3 |

9

10 **Table S2.** List of hit numbers for the approved drugs.

| SS | AD      |         | GPCR    |         | Ion Channel |         | Protease |         | Kinase |         |
|----|---------|---------|---------|---------|-------------|---------|----------|---------|--------|---------|
|    | Ave     | Max     | Ave     | Max     | Ave         | Max     | Ave      | Max     | Ave    | Max     |
| 85 | 60.9    | 387.4   | 54.1    | 267.9   | 82.8        | 337.3   | 64.1     | 392.7   | 22.5   | 54      |
| 80 | 120.8   | 740.6   | 98.1    | 380.3   | 184.2       | 600.5   | 103      | 555.3   | 47.6   | 117.8   |
| 75 | 247.6   | 1328.7  | 224.5   | 666.5   | 425.6       | 1166.6  | 180      | 733.3   | 113.7  | 328.1   |
| 70 | 579.1   | 2485.3  | 589.7   | 1546    | 1128.4      | 2983.2  | 430.3    | 1336.6  | 284.8  | 967.5   |
| 65 | 1426.4  | 4895.6  | 1619.4  | 3983.1  | 2883.3      | 7181.4  | 1150     | 3185.4  | 695.5  | 2574.7  |
| 60 | 3442.8  | 9767.7  | 4251.2  | 9054.4  | 6816.7      | 14978.6 | 3145     | 7995.7  | 1671   | 6275.5  |
| 55 | 8576.9  | 20773.3 | 11282.6 | 20736.1 | 16074.4     | 31261.6 | 8723.9   | 20363.5 | 4083.9 | 14940.3 |
| 50 | 20459.8 | 43533.8 | 28221.5 | 44412   | 35923.7     | 61691.1 | 22467.4  | 49053.1 | 9482.8 | 32063.7 |

11

12 **Table S3.** List of hit numbers for the investigational drugs.

| SS | ID  |     | GPCR |     | Ion Channel |     | Protease |     | Kinase |     |
|----|-----|-----|------|-----|-------------|-----|----------|-----|--------|-----|
|    | Ave | Max | Ave  | Max | Ave         | Max | Ave      | Max | Ave    | Max |

## Supporting Information

|           |        |         |         |         |         |         |         |         |         |         |
|-----------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| <b>85</b> | 59     | 336.4   | 42.7    | 167.4   | 159.9   | 476.3   | 24.6    | 84.7    | 22.3    | 49.7    |
| <b>80</b> | 105.7  | 460.5   | 92.9    | 260.8   | 294.1   | 739.3   | 54.8    | 139.9   | 48.7    | 101.7   |
| <b>75</b> | 206    | 686.9   | 259     | 675.1   | 525.9   | 1156    | 121.5   | 278.1   | 126.4   | 324.4   |
| <b>70</b> | 475.2  | 1409.7  | 746.3   | 1950.7  | 1065.5  | 2307.5  | 341.3   | 680     | 344.1   | 1088.2  |
| <b>65</b> | 1204   | 3487.7  | 2058.7  | 5017.8  | 2431    | 5410.8  | 1102.2  | 2050    | 865.2   | 3118.5  |
| <b>60</b> | 3106.6 | 8518.9  | 5483.9  | 12233   | 5830.7  | 11808.5 | 3503.3  | 6283    | 2051.7  | 7906    |
| <b>55</b> | 8217.7 | 20849.2 | 14613.5 | 29586.9 | 13824.4 | 25527   | 10808.6 | 17514.4 | 4773.6  | 18957.3 |
| <b>50</b> | 20642  | 47691.7 | 36838.6 | 66926.3 | 32014.8 | 53729.6 | 29357.2 | 44219.4 | 10509.8 | 40906.7 |

13

14 **Table S4.** List of hit numbers for the experimental drugs.

| <b>SS</b> | <b>ED</b> |         | <b>GPCR</b> |         | <b>Ion Channel</b> |         | <b>Protease</b> |         | <b>Kinase</b> |         |
|-----------|-----------|---------|-------------|---------|--------------------|---------|-----------------|---------|---------------|---------|
|           | Ave       | Max     | Ave         | Max     | Ave                | Max     | Ave             | Max     | Ave           | Max     |
| <b>85</b> | 56.7      | 421.6   | 242.6       | 783.9   | 112.8              | 613.7   | 64.8            | 436.8   | 13.8          | 178.3   |
| <b>80</b> | 110.6     | 757.4   | 359.6       | 975     | 176.7              | 769.1   | 144.9           | 780.6   | 25            | 234.1   |
| <b>75</b> | 225.5     | 1322.1  | 698.9       | 1536.9  | 329                | 1124.9  | 343.7           | 1605.8  | 50.8          | 342.8   |
| <b>70</b> | 503.5     | 2447.4  | 1436.9      | 3215.2  | 746.2              | 2364    | 861.9           | 3368.5  | 114.8         | 617.1   |
| <b>65</b> | 1172.2    | 4656.1  | 3203.4      | 7641.9  | 1761               | 5589.3  | 2172.7          | 6925.9  | 274.8         | 1117.9  |
| <b>60</b> | 2762.9    | 9397.4  | 7294.4      | 17724.8 | 3899               | 12181   | 5421.1          | 14722.2 | 703.4         | 2227.2  |
| <b>55</b> | 6862.1    | 20258.4 | 17314.7     | 41747.4 | 9635.9             | 28598.8 | 13769.9         | 31913.9 | 2064          | 5445.6  |
| <b>50</b> | 16376.3   | 42727.1 | 38503.8     | 88916.5 | 21525.5            | 60150.3 | 33073.7         | 66725.7 | 6137.1        | 14399.4 |

15

16

**Table S5.** Percentages of drugs having zero hits

| <b>SS</b> | <b>Type</b> | <b>All</b> | <b>GPCR</b> | <b>Ion Channel</b> | <b>Protease</b> | <b>Kinase</b> |
|-----------|-------------|------------|-------------|--------------------|-----------------|---------------|
| <b>85</b> | All         | 54.4       | 19.7        | 30.2               | 65.4            | 61.6          |
|           | AD          | 36.4       | 17.5        | 20.6               | 64.8            | 38.4          |
|           | ID          | 39.1       | 24.4        | 18.2               | 55.6            | 37.2          |
|           | ED          | 61.4       | 34.3        | 55.3               | 65.6            | 65.7          |
| <b>80</b> | All         | 47.3       | 17          | 25.3               | 55.9            | 53.2          |
|           | AD          | 31.2       | 15.4        | 15.9               | 59.2            | 35.4          |
|           | ID          | 35.2       | 20.7        | 14.5               | 50              | 34.9          |
|           | ED          | 53.6       | 31.4        | 50.5               | 55.5            | 56.4          |
| <b>75</b> | All         | 40.8       | 14          | 19.4               | 45.9            | 45.6          |
|           | AD          | 27.4       | 12.7        | 10.5               | 56.8            | 32.3          |
|           | ID          | 31.7       | 18.5        | 12.7               | 50              | 30.2          |
|           | ED          | 46         | 28.6        | 41.7               | 44.4            | 48.3          |
| <b>70</b> | All         | 32.9       | 10.9        | 14.7               | 35.6            | 36            |
|           | AD          | 22.6       | 9.3         | 6.9                | 51.2            | 26.3          |
|           | ID          | 26.5       | 14.1        | 12.7               | 44.4            | 25.6          |
|           | ED          | 36.8       | 25.7        | 33                 | 33.2            | 37.9          |
| <b>65</b> | All         | 26         | 7           | 12                 | 26.9            | 25.6          |
|           | AD          | 18.1       | 5.7         | 5.1                | 45.6            | 19.2          |

## Supporting Information

|           |     |      |      |      |      |      |
|-----------|-----|------|------|------|------|------|
|           | ID  | 21   | 8.1  | 10.9 | 36.1 | 20.9 |
|           | ED  | 29   | 25.7 | 28.2 | 24.1 | 26.7 |
| <b>60</b> | All | 19.5 | 5    | 9.1  | 20.3 | 16.1 |
|           | AD  | 14.7 | 4.2  | 3.2  | 44   | 16.2 |
|           | ID  | 15.4 | 6.7  | 9.1  | 27.8 | 14   |
|           | ED  | 21.5 | 20   | 22.3 | 16.9 | 16.3 |
| <b>55</b> | All | 13.9 | 2.7  | 5.7  | 15   | 10.8 |
|           | AD  | 10.7 | 2.1  | 1.8  | 42.4 | 9.1  |
|           | ID  | 9.9  | 3    | 7.3  | 19.4 | 4.7  |
|           | ED  | 15.2 | 14.3 | 13.6 | 11.4 | 11.2 |
| <b>50</b> | All | 10.1 | 1.4  | 2.7  | 10.5 | 6.6  |
|           | AD  | 8.1  | 1.1  | 1.1  | 32.8 | 5.1  |
|           | ID  | 6.4  | 0.7  | 0    | 16.7 | 4.7  |
|           | ED  | 10.9 | 8.6  | 6.8  | 7.5  | 6.8  |

17

18 **Table S6.** List of entries of actives collected in ZINC database. Entries that are duplicated with drug  
19 molecules were removed.

| Active Set  | # Entries | Remark   |
|-------------|-----------|--|
| All         | 5847      | Collected in the ZINC target subset  |
| All_Human   | 4078      | All entries for testing on human drug targets  |
| GPCR        | 96        | Non-redundant entries for Cannabinoid Receptor 1, Cannabinoid Receptor 2, Neurokinin 1 Receptor, Neurokinin 2 Receptor and Neurokinin 3 Receptor |
| Ion Channel | 50        | Non-redundant entries for voltage-gated calcium channel Alpha2/delta, voltage-gated potassium channel subunit Kv1.3, and HERG                    |
| Kinase      | 55        | Tyrosine protein kinase, Adenosine kinase and Serine/threonine protein kinase  |
| Protease    | 4         | HIV-1 Protease, Cysteine Protease Falcipain-3, Chymase, Human Rhinovirus A protease and hepatitis C virus NS3 protease                           |

20

21 **Table S7.** Performance of druglike datasets against the "All\_Human" active set which has 4078 actives  
22 tested on human drug targets.

| SS        | Drug  |      |      | AD    |      |      | AD_ID |      |      | ED    |      |      |
|-----------|-------|------|------|-------|------|------|-------|------|------|-------|------|------|
|           | #Hits | HR   | EF   | #Hits | HR   | EF   | #Hits | HR   | EF   | #Hits | HR   | EF   |
| <b>85</b> | 979   | 24.0 | 36.8 | 563   | 13.8 | 51.9 | 604   | 14.8 | 54.2 | 527   | 12.9 | 24.6 |
| <b>80</b> | 1167  | 28.6 | 21.1 | 658   | 16.1 | 32.0 | 706   | 17.3 | 33.0 | 677   | 16.6 | 15.3 |
| <b>75</b> | 1417  | 34.7 | 12.4 | 803   | 19.7 | 18.6 | 865   | 21.2 | 19.1 | 883   | 21.7 | 10.0 |
| <b>70</b> | 1693  | 41.5 | 7.1  | 957   | 23.5 | 9.9  | 1018  | 25.0 | 9.8  | 1111  | 27.2 | 6.2  |
| <b>65</b> | 2033  | 49.9 | 4.3  | 1169  | 28.7 | 5.6  | 1247  | 30.6 | 5.5  | 1439  | 35.3 | 4.0  |
| <b>60</b> | 2496  | 61.2 | 2.9  | 1506  | 36.9 | 3.6  | 1595  | 39.1 | 3.5  | 1897  | 46.5 | 2.8  |
| <b>55</b> | 2933  | 71.9 | 2.1  | 1897  | 46.5 | 2.5  | 1987  | 48.7 | 2.4  | 2446  | 60.0 | 2.1  |
| <b>50</b> | 3311  | 81.2 | 1.6  | 2335  | 57.3 | 1.8  | 2431  | 59.6 | 1.8  | 2984  | 73.2 | 1.6  |

23

## Supporting Information

24 **Table S8.** Performance of druglike datasets against the “GPCR” external active set which has 96 entries.

| SS | Drug  |      |     | AD    |      |     | AD_ID |      |     | ED    |      |     | GPCR  |      |      |
|----|-------|------|-----|-------|------|-----|-------|------|-----|-------|------|-----|-------|------|------|
|    | #Hits | HR   | EF  | #Hits | HR   | EF  | #Hits | HR   | EF  | #Hits | HR   | EF  | #Hits | HR   | EF   |
| 85 | 2     | 2.1  | 3.2 | 2     | 2.1  | 7.8 | 2     | 2.1  | 7.6 | 0     | 0.0  | 0.0 | 2     | 2.1  | 15.4 |
| 80 | 6     | 6.3  | 4.6 | 2     | 2.1  | 4.1 | 2     | 2.1  | 4.0 | 4     | 4.2  | 3.8 | 2     | 2.1  | 8.8  |
| 75 | 10    | 10.4 | 3.7 | 2     | 2.1  | 2.0 | 3     | 3.1  | 2.8 | 7     | 7.3  | 3.4 | 2     | 2.1  | 4.1  |
| 70 | 14    | 14.6 | 2.5 | 3     | 3.1  | 1.3 | 4     | 4.2  | 1.6 | 10    | 10.4 | 2.4 | 2     | 2.1  | 1.7  |
| 65 | 22    | 22.9 | 2.0 | 9     | 9.4  | 1.8 | 14    | 14.6 | 2.6 | 16    | 16.7 | 1.9 | 6     | 6.3  | 2.1  |
| 60 | 45    | 46.9 | 2.2 | 20    | 20.8 | 2.0 | 24    | 25.0 | 2.2 | 35    | 36.5 | 2.2 | 19    | 19.8 | 3.0  |
| 55 | 82    | 85.4 | 2.4 | 52    | 54.2 | 2.9 | 56    | 58.3 | 2.8 | 73    | 76.0 | 2.6 | 36    | 37.5 | 2.8  |
| 50 | 87    | 90.6 | 1.8 | 65    | 67.7 | 2.2 | 70    | 72.9 | 2.2 | 82    | 85.4 | 1.9 | 53    | 55.2 | 2.4  |

25

26 **Table S9.** Performance of druglike datasets against the “Ion Channel” external active set which has 50  
27 entries.

| SS | Drug  |      |      | AD    |      |       | AD_ID |      |       | ED    |      |      | ION   |      |       |
|----|-------|------|------|-------|------|-------|-------|------|-------|-------|------|------|-------|------|-------|
|    | #Hits | HR   | EF   | #Hits | HR   | EF    | #Hits | HR   | EF    | #Hits | HR   | EF   | #Hits | HR   | EF    |
| 85 | 28    | 56.0 | 85.8 | 27    | 54.0 | 202.9 | 27    | 54.0 | 197.4 | 6     | 12.0 | 22.9 | 17    | 34.0 | 291.6 |
| 80 | 30    | 60.0 | 44.2 | 29    | 58.0 | 115.1 | 29    | 58.0 | 110.5 | 6     | 12.0 | 11.1 | 17    | 34.0 | 152.3 |
| 75 | 30    | 60.0 | 21.4 | 29    | 58.0 | 54.9  | 29    | 58.0 | 52.1  | 6     | 12.0 | 5.6  | 19    | 38.0 | 75.9  |
| 70 | 31    | 62.0 | 10.7 | 30    | 60.0 | 25.2  | 30    | 60.0 | 23.6  | 6     | 12.0 | 2.7  | 19    | 38.0 | 30.8  |
| 65 | 31    | 62.0 | 5.4  | 30    | 60.0 | 11.7  | 30    | 60.0 | 10.8  | 7     | 14.0 | 1.6  | 19    | 38.0 | 13.1  |
| 60 | 34    | 68.0 | 3.2  | 33    | 66.0 | 6.5   | 33    | 66.0 | 5.9   | 10    | 20.0 | 1.2  | 23    | 46.0 | 7.4   |
| 55 | 39    | 78.0 | 2.2  | 34    | 68.0 | 3.6   | 36    | 72.0 | 3.5   | 19    | 38.0 | 1.3  | 24    | 48.0 | 3.8   |
| 50 | 46    | 92.0 | 1.8  | 38    | 76.0 | 2.4   | 40    | 80.0 | 2.4   | 32    | 64.0 | 1.4  | 26    | 52.0 | 2.3   |

28

29 **Table S10.** Performance of druglike datasets against the “Kinase” external active set which has 55 entries.

| SS | Drug  |      |      | AD    |      |      | AD_ID |      |      | ED    |      |      | KINASE |      |       |
|----|-------|------|------|-------|------|------|-------|------|------|-------|------|------|--------|------|-------|
|    | #Hits | HR   | EF   | #Hits | HR   | EF   | #Hits | HR   | EF   | #Hits | HR   | EF   | #Hits  | HR   | EF    |
| 85 | 16    | 29.1 | 44.6 | 6     | 10.9 | 41.0 | 6     | 10.9 | 39.9 | 14    | 25.5 | 48.5 | 14     | 25.5 | 501.7 |
| 80 | 18    | 32.7 | 24.1 | 7     | 12.7 | 25.3 | 7     | 12.7 | 24.3 | 16    | 29.1 | 26.8 | 15     | 27.3 | 291.1 |
| 75 | 21    | 38.2 | 13.6 | 8     | 14.5 | 13.8 | 8     | 14.5 | 13.1 | 19    | 34.5 | 16.0 | 18     | 32.7 | 169.3 |
| 70 | 23    | 41.8 | 7.2  | 9     | 16.4 | 6.9  | 9     | 16.4 | 6.4  | 21    | 38.2 | 8.6  | 21     | 38.2 | 86.0  |
| 65 | 26    | 47.3 | 4.1  | 13    | 23.6 | 4.6  | 13    | 23.6 | 4.2  | 25    | 45.5 | 5.2  | 23     | 41.8 | 40.0  |
| 60 | 30    | 54.5 | 2.6  | 16    | 29.1 | 2.9  | 16    | 29.1 | 2.6  | 29    | 52.7 | 3.2  | 24     | 43.6 | 17.6  |
| 55 | 31    | 56.4 | 1.6  | 17    | 30.9 | 1.6  | 18    | 32.7 | 1.6  | 29    | 52.7 | 1.8  | 26     | 47.3 | 7.7   |
| 50 | 37    | 67.3 | 1.3  | 24    | 43.6 | 1.4  | 24    | 43.6 | 1.3  | 37    | 67.3 | 1.5  | 32     | 58.2 | 4.1   |

30

31 **Table S11.** Performance of druglike datasets against the “Protease” external active set which has 4 entries.

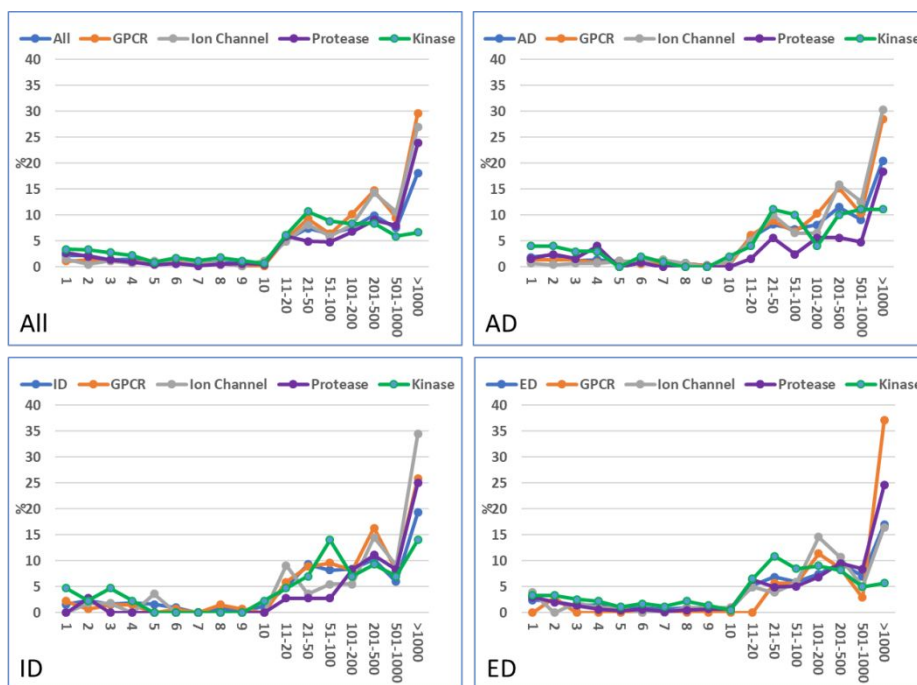
| SS | Drug  |    |    | AD    |    |    | AD_ID |    |    | ED    |    |    | PROTEASE |    |    |
|----|-------|----|----|-------|----|----|-------|----|----|-------|----|----|----------|----|----|
|    | #Hits | HR | EF | #Hits | HR | EF | #Hits | HR | EF | #Hits | HR | EF | #Hits    | HR | EF |

## Supporting Information

|    |   |      |       |   |      |       |   |      |       |   |      |     |   |      |     |
|----|---|------|-------|---|------|-------|---|------|-------|---|------|-----|---|------|-----|
| 85 | 3 | 75.0 | 114.9 | 3 | 75.0 | 281.8 | 3 | 75.0 | 274.2 | 0 | 0.0  | 0.0 | 0 | 0.0  | 0.0 |
| 80 | 3 | 75.0 | 55.3  | 3 | 75.0 | 148.9 | 3 | 75.0 | 142.9 | 0 | 0.0  | 0.0 | 0 | 0.0  | 0.0 |
| 75 | 3 | 75.0 | 26.7  | 3 | 75.0 | 71.0  | 3 | 75.0 | 67.4  | 0 | 0.0  | 0.0 | 0 | 0.0  | 0.0 |
| 70 | 3 | 75.0 | 12.9  | 3 | 75.0 | 31.5  | 3 | 75.0 | 29.5  | 0 | 0.0  | 0.0 | 0 | 0.0  | 0.0 |
| 65 | 3 | 75.0 | 6.5   | 3 | 75.0 | 14.6  | 3 | 75.0 | 13.5  | 0 | 0.0  | 0.0 | 0 | 0.0  | 0.0 |
| 60 | 3 | 75.0 | 3.6   | 3 | 75.0 | 7.4   | 3 | 75.0 | 6.7   | 1 | 25.0 | 1.5 | 1 | 25.0 | 2.8 |
| 55 | 3 | 75.0 | 2.1   | 3 | 75.0 | 4.0   | 3 | 75.0 | 3.6   | 1 | 25.0 | 0.9 | 1 | 25.0 | 1.4 |
| 50 | 3 | 75.0 | 1.5   | 3 | 75.0 | 2.4   | 3 | 75.0 | 2.3   | 1 | 25.0 | 0.6 | 1 | 25.0 | 0.8 |

32

33

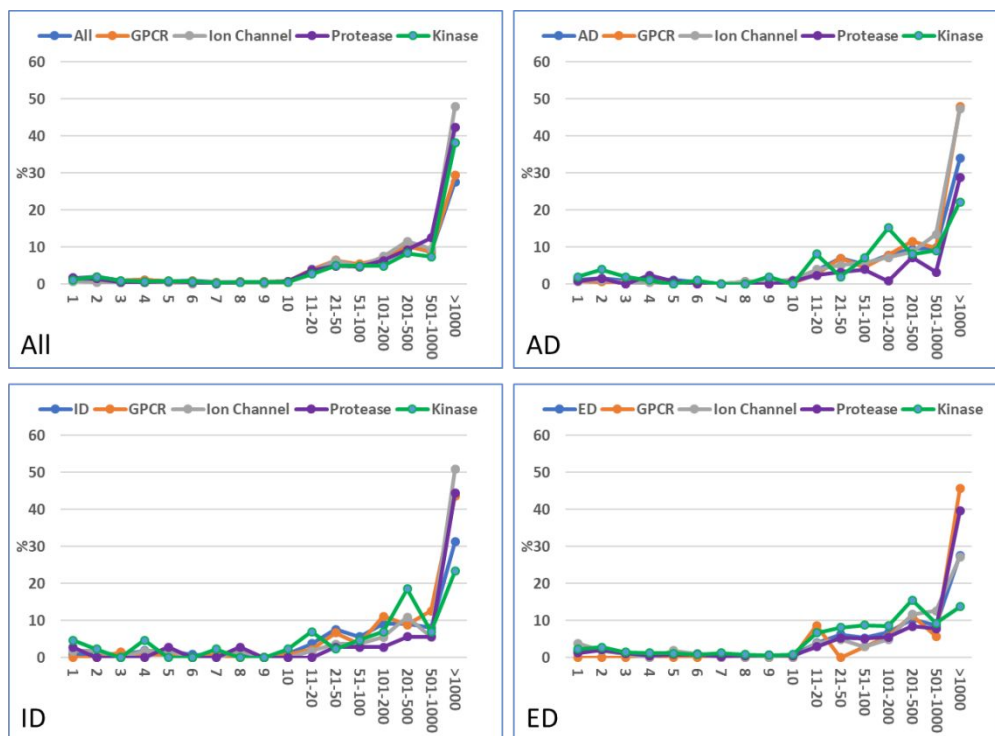


34

35 **Figure S1.** Distributions of drug molecules (%) in 17 hit number groups. A molecule of the ZINC  
 36 druglike dataset is recognized as a hit if its 2D-similarity score against the query drug molecule  
 37 is equal to or better than 0.65 (SS65).

38

## Supporting Information



39

40 **Figure S2.** Distributions of drug molecules (%) in 17 hit number groups. A molecule of the ZINC  
41 druglike dataset is recognized as a hit if its 2D-similarity score against the query drug molecule  
42 is equal to or better than 0.60 (SS60).

43

## Supporting Information

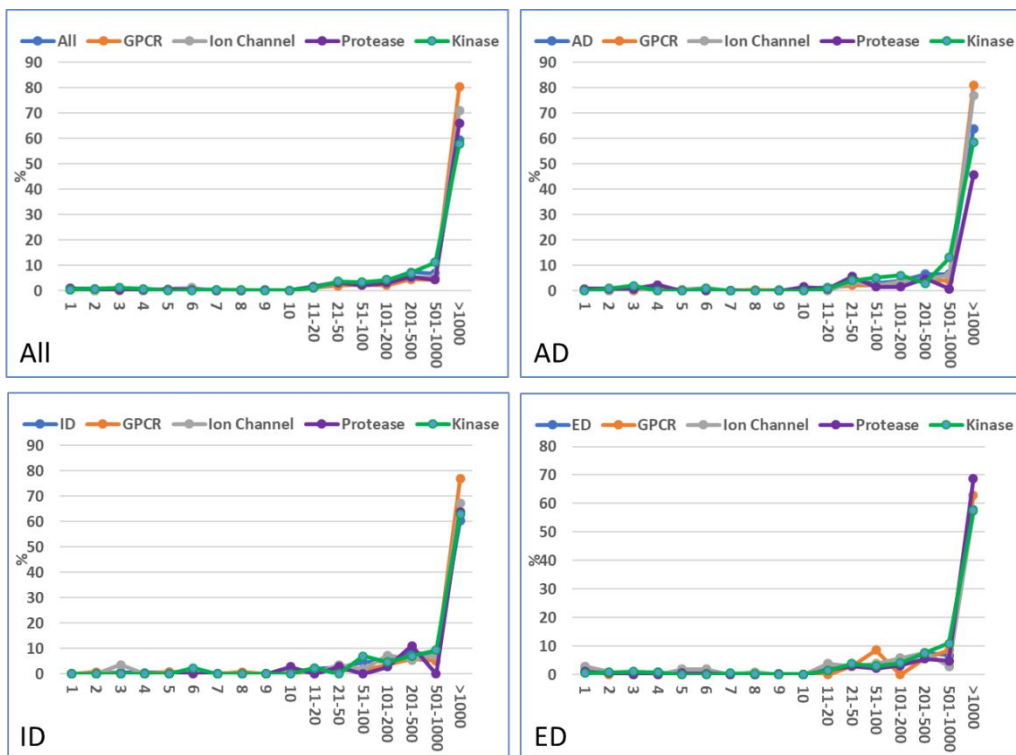


44

45 **Figure S3.** Distributions of drug molecules (%) in 17 hit number groups. A molecule of the ZINC  
46 druglike dataset is recognized as a hit if its 2D-similarity score against the query drug molecule  
47 is equal to or better than 0.55 (SS55).

48

## Supporting Information



49

50 **Figure S4.** Distributions of drug molecules (%) in 17 hit number groups. A molecule of the ZINC  
51 druglike dataset is recognized as a hit if its 2D-similarity score against the query drug molecule  
52 is equal to or better than 0.50 (SS50).