# Supplementary material

**Table S1**. Number and percentage of compounds already present in the dataset of the models.

|  |  |  |  |
| --- | --- | --- | --- |
|  | % of experimental data on 628 negative compounds (number of not concordant values) | % of experimental data on 97 ‘cleaned’ positive compounds (number of not concordant values) | % of experimental data on 74 additional positive compounds (number of not concordant values) |
| CAESAR | 9.1% (6) | 38.1% (5) | 8.1% (0) |
| ISS | 6.8% (12) | 47.4% (2) | 5.4% (0) |
| Antares | 11.9% (6) | 52.6% (6) | 14.9% (2) |
| ISSCAN-CGX | 8.9% (13) | 53.6% (2) | 9.5% (0) |
| OSF | 11.5% (5) | 46.4% (9) | 10.8% (0) |

**Table S2**. Number and percentage of compounds already present in the dataset of the models.

|  |  |  |
| --- | --- | --- |
|  | Number of carcinogenic predictions for the 26 non genotoxic carcinogens in the ‘cleaned’ dataset (n = 94 positive) | Number of carcinogenic predictions for the 13 non genotoxic carcinogens among the 74 additional positive compounds |
| CAESAR | 7 | 11 |
| ISS | 16 | 6 |
| Antares | 13 | 0 |
| ISSCAN-CGX | 14 | 0 |
| OSF | 12 | 11 |