External Validation of the ACD/Percepta Genotoxicity model

1) Performance statistics of the validation set

Table 1. Model performance using a probability threshold of $p = 0.5$ and considering an RI value > 0.3.

	Calculated probability (*p*)			
Observed	**>0.5**	**<0.5**	**N_{obj}**	**Performance**
Genotoxic	928	67	995	Sensitivity = 93.3%
Safe	96	392	488	Specificity = 80.3%
			1483	Accuracy = 89.0%

Table 2. Model performance using a probability threshold of $p = 0.5$ and considering an RI value > 0.5.

	Calculated probability (*p*)			
Observed	**>0.5**	**<0.5**	**N_{obj}**	**Performance**
Genotoxic	786	23	809	Sensitivity = 97.2%
Safe	51	257	308	Specificity = 83.4%
			1117	Accuracy = 93.4%

Table 3. Model performance using the probability thresholds $p<0.2$, $0.2<p<0.8$ and $p>0.8$, and considering an RI value > 0.3.

	Calculated probability (*p*)				
Observed	**>0.8**	**0.2-0.8**	**<0.2**	**N_{obj}***	**Performance**
Genotoxic	848	125	22	870	Sensitivity = 97.5%
Safe	48	130	310	358	Specificity = 86.6%
				1228	Accuracy = 94.3%

* inconclusive predictions (0.2<p<0.8) were not used for performance statistics.

Table 4. Model performance using the probability thresholds $p<0.2$, $0.2<p<0.8$ and $p>0.8$, and considering an RI value > 0.5.

	Calculated probability (*p*)				
Observed	**>0.8**	**0.2-0.8**	**<0.2**	**N_{obj}***	**Performance**
Genotoxic	760	38	11	771	Sensitivity = 98.6%
Safe	32	51	225	257	Specificity = 87.5%
				1028	Accuracy = 95.8%

* inconclusive predictions (0.2<p<0.8) were not used for performance statistics.

2) Performance statistics of an additional validation set (Case study)

External predicting power of the ACD/Percepta Genotoxicity model was tested on the drug-like PDR 1999-2008 dataset, obtained from a publication by Snyder (*Environ. Mol. Mutagen*. 2009; 50(6): 435-50).

Table 5. Model performance using a probability threshold of $p = 0.5$ and considering an RI value > 0.3.

	Calculated probability (*p*)			
Observed	**≥0.5**	**<0.5**	**N_{obj}**	**Performance**
Ames positive	30	9	39	Sensitivity = 76.9%
Ames negative	11	422	433	Specificity = 97.5%
			472	Accuracy = 95.7%

Table 6. Model performance using the probability thresholds $p \leq 0.3$, $0.3 < p < 0.7$ and $p \geq 0.7$, and considering an RI value > 0.3.

	Calculated probability (*p*)				
Observed	**≥0.7**	**0.3-0.7**	**≤0.3**	**N_{obj}***	**Performance**
Ames positive	20	14	5	25	Sensitivity = 80.0%
Ames negative	3	32	398	401	Specificity = 99.3%
				426	Accuracy = 98.1%

* inconclusive predictions (0.3<p<0.7) were not used for performance statistics.