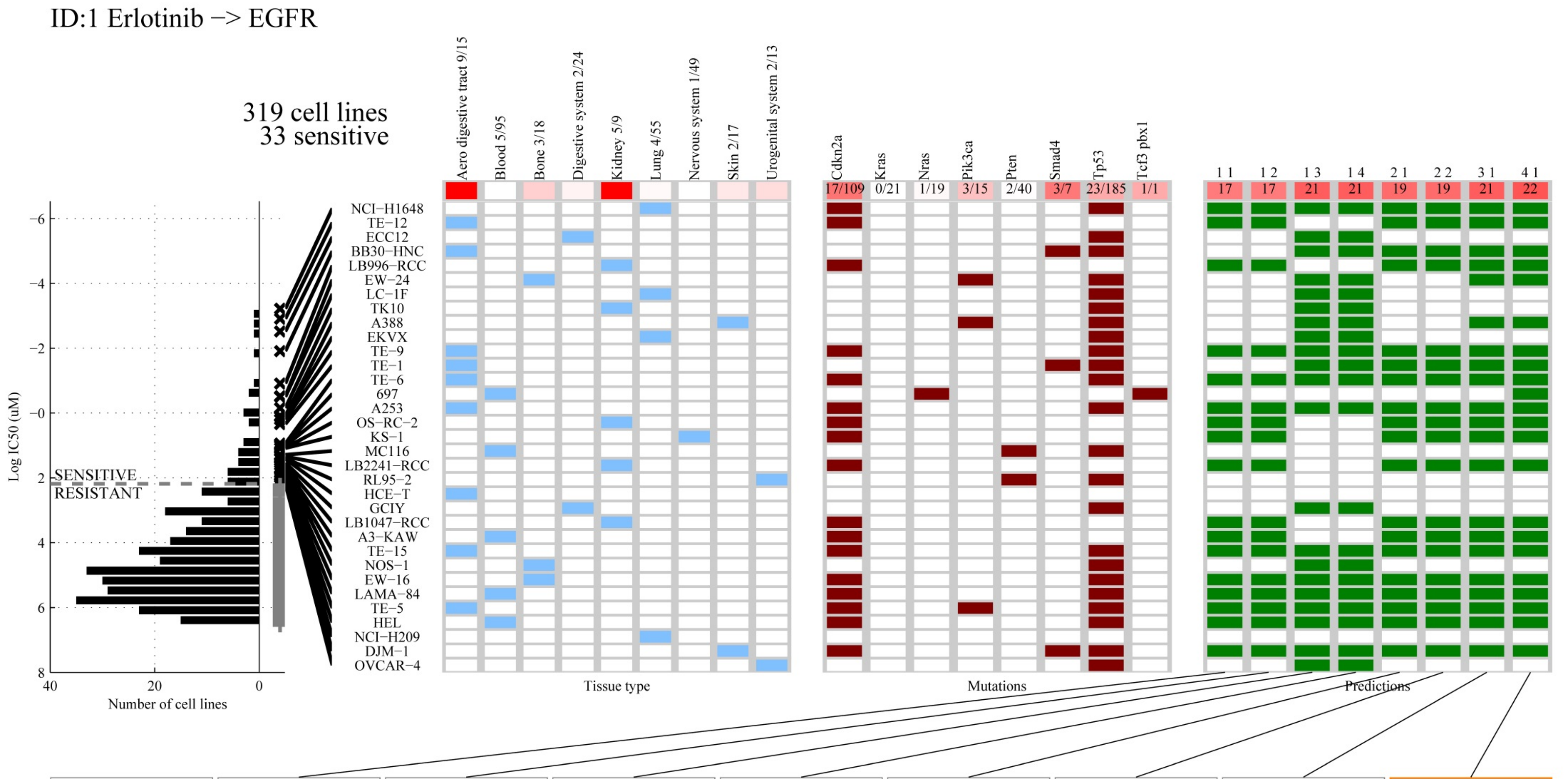
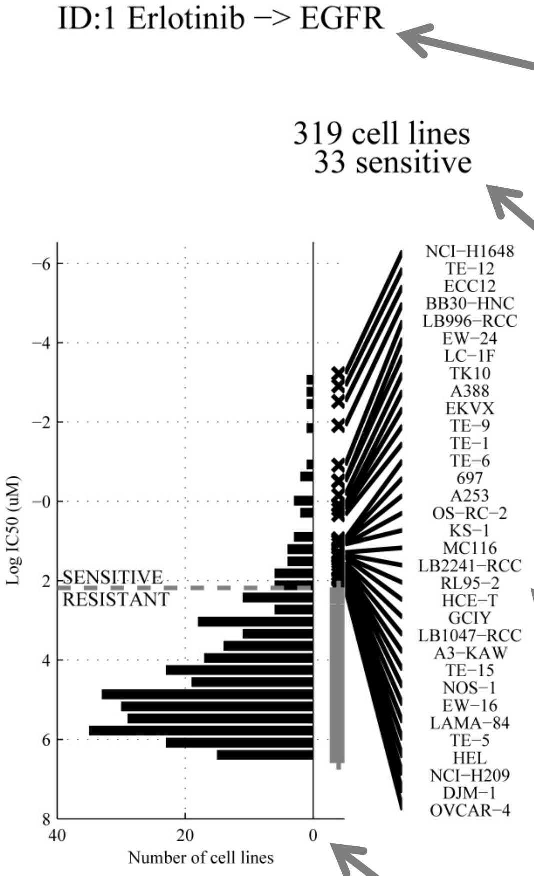
Visual explanation of the logic models for a single drug inferred by LOBICO



Model name K M	1 1 1 1	1 2 1 2	1 3 1 3	1 4 1 4	2 1 2 1	2 2 2 2	3 1 3 1	4 1 4 1
Logic formula	**CDKN2**	**CDKN2&~PTEN**	**~KRAS&~PTEN& TP53**	**~KRAS&~NRAS& ~PTEN& TP53**	**CDKN2 | SMAD4**	**[~KRAS&SMAD4] | [CDKN2&~PTEN]**	**CDKN2 | PIK3C | SMAD4**	**CDKN2 | PIK3C | SMAD4 | TCF3**
TP FP FN TN	17 92 16 194	17 75 16 211	21 120 12 166	21 113 12 173	19 95 14 191	19 77 14 209	21 104 12 182	22 104 11 182
Specificity Precision Recall	0.68 0.16 0.52	0.74 0.18 0.52	0.58 0.15 0.64	0.6 0.16 0.64	0.67 0.17 0.58	0.73 0.2 0.58	0.64 0.17 0.64	0.64 0.17 0.67

Information about the drug and the cell lines sensitive to the drug



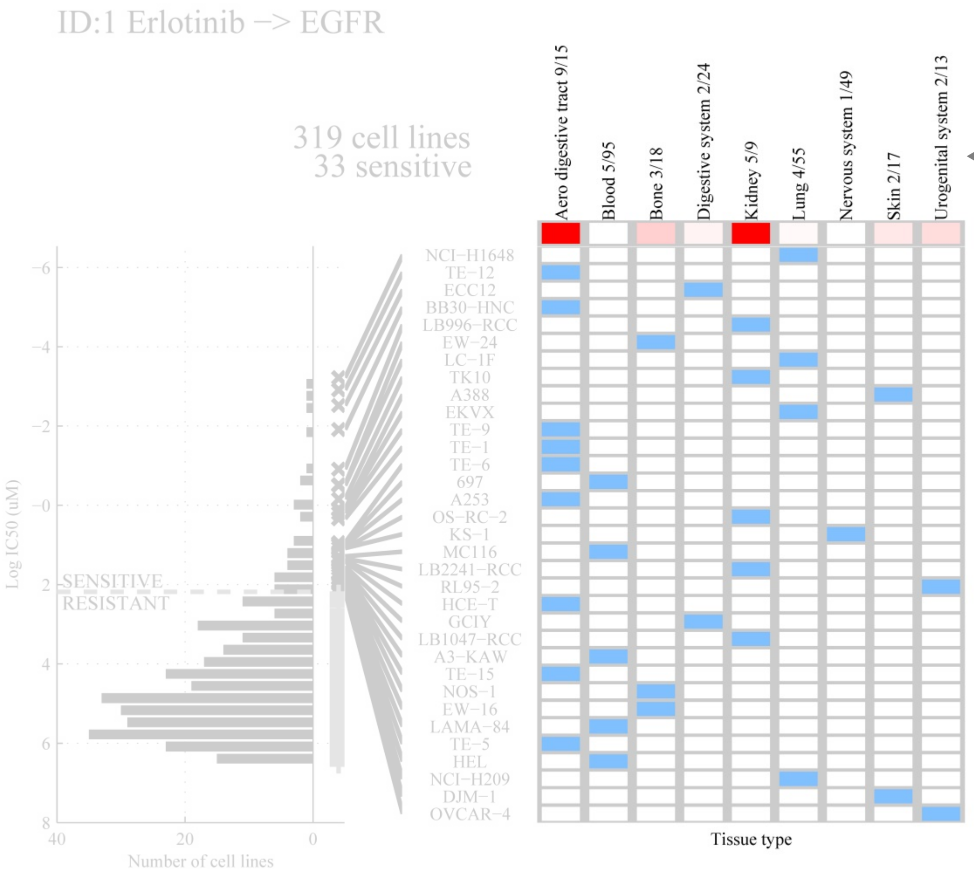
Drug ID followed by the name of the drug (if available) and, to the right of ‘->’ the target(s) of the drug (if available).

Number of cell lines for which drug response values (IC50s) are available, and below, the number of cell lines that are sensitive to the drug.

Names of the sensitive cell lines with the most sensitive (lowest IC50) at the top.

Histogram of all (log transformed) IC50s. The grey dashed line represents the threshold used to decide whether cell lines are sensitive or resistant.

Tissue of origin of the sensitive cell lines



Tissue type labels followed by ‘*x*/*y*’, where *x* is the number of sensitive cell lines with the indicated tissue type, and *y* is the total number of cell lines from that tissue type

White to red heatmap indicating whether cell lines from the indicated tissue type are enriched within the group of sensitive cell lines. Bright red indicates strong enrichment. White indicates no enrichment.

Example: Of the 33 sensitive cell lines, 5 are ‘kidney’ lines (5th column), which is a significant enrichment given that there are 9 ‘kidney’ lines in the total set of 319 cell lines.

Binary heatmap indicating for each of the sensitive cell lines the tissue of origin (indicated by light blue).

Example: The most sensitive cell line (NCI-H1648, first row) is a ‘lung’ cell line.

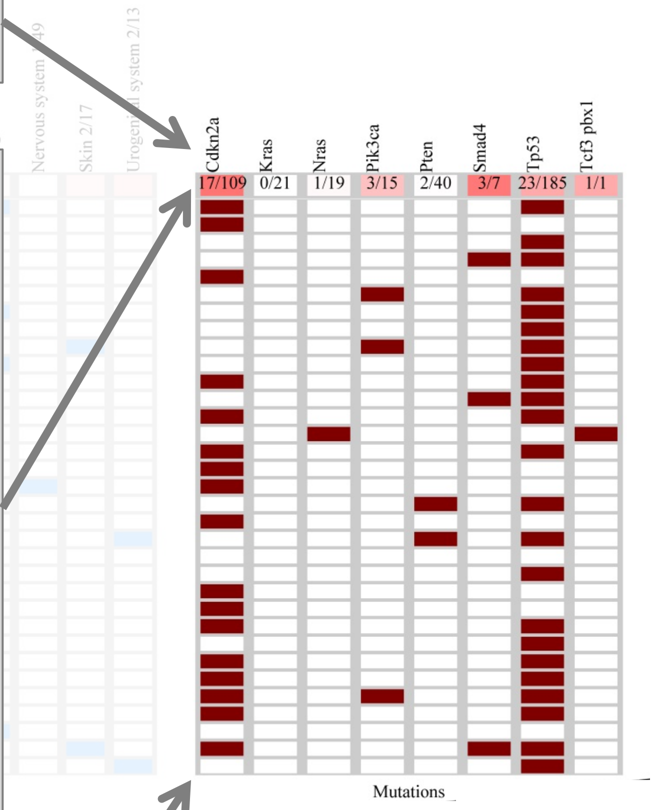
Mutation status of the sensitive cell lines

Labels of the genomic features used in one or more of the inferred logic models

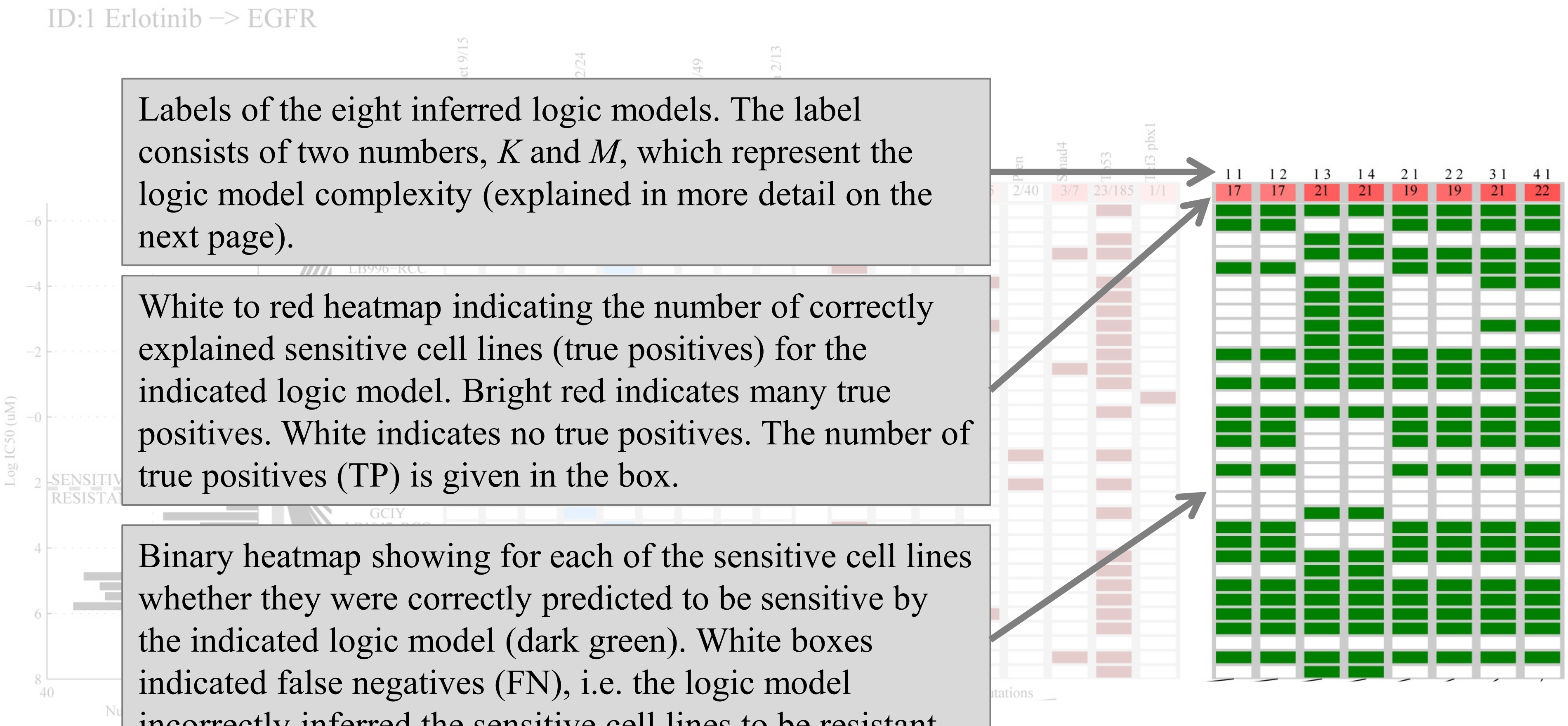
White to red heatmap indicating whether cell lines that have a mutation of the indicated genomic feature are enriched within the group of sensitive cell lines. Bright red indicates strong enrichment. White indicates no enrichment. The text in the box is ‘*x*/*y*’, where *x* is the number of mutated sensitive cell lines, and *y* is the total number of cell lines with a mutation of that feature.

Example: Of the 33 sensitive cell lines, 3 have a ‘Smad4’ mutation (6th column), which is a significant enrichment given that there are 7 lines with a ‘Smad4’ mutation in the total set of 319 cell lines.

Binary heatmap showing for each of the sensitive cell lines whether they have a mutation of the indicated feature (dark red).



Correctly and incorrectly predicted sensitive cell lines

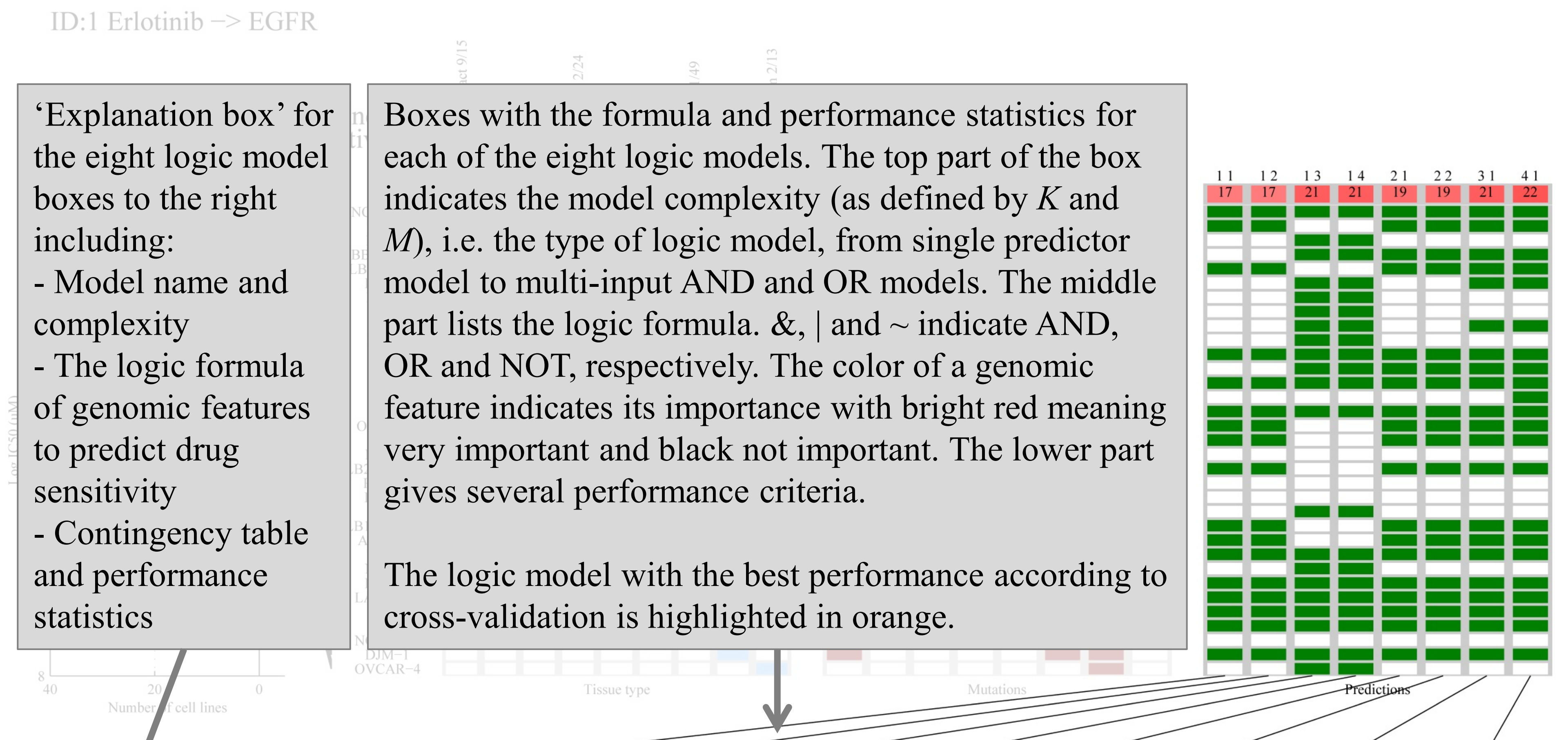


Labels of the eight inferred logic models. The label consists of two numbers, *K* and *M*, which represent the logic model complexity (explained in more detail on the next page).

White to red heatmap indicating the number of correctly explained sensitive cell lines (true positives) for the indicated logic model. Bright red indicates many true positives. White indicates no true positives. The number of true positives (TP) is given in the box.

Binary heatmap showing for each of the sensitive cell lines whether they were correctly predicted to be sensitive by the indicated logic model (dark green). White boxes indicated false negatives (FN), i.e. the logic model incorrectly inferred the sensitive cell lines to be resistant.

Formulas and performance statistics for the logic models



‘Explanation box’ for the eight logic model boxes to the right including:
- Model name and complexity
- The logic formula of genomic features to predict drug sensitivity
- Contingency table and performance statistics

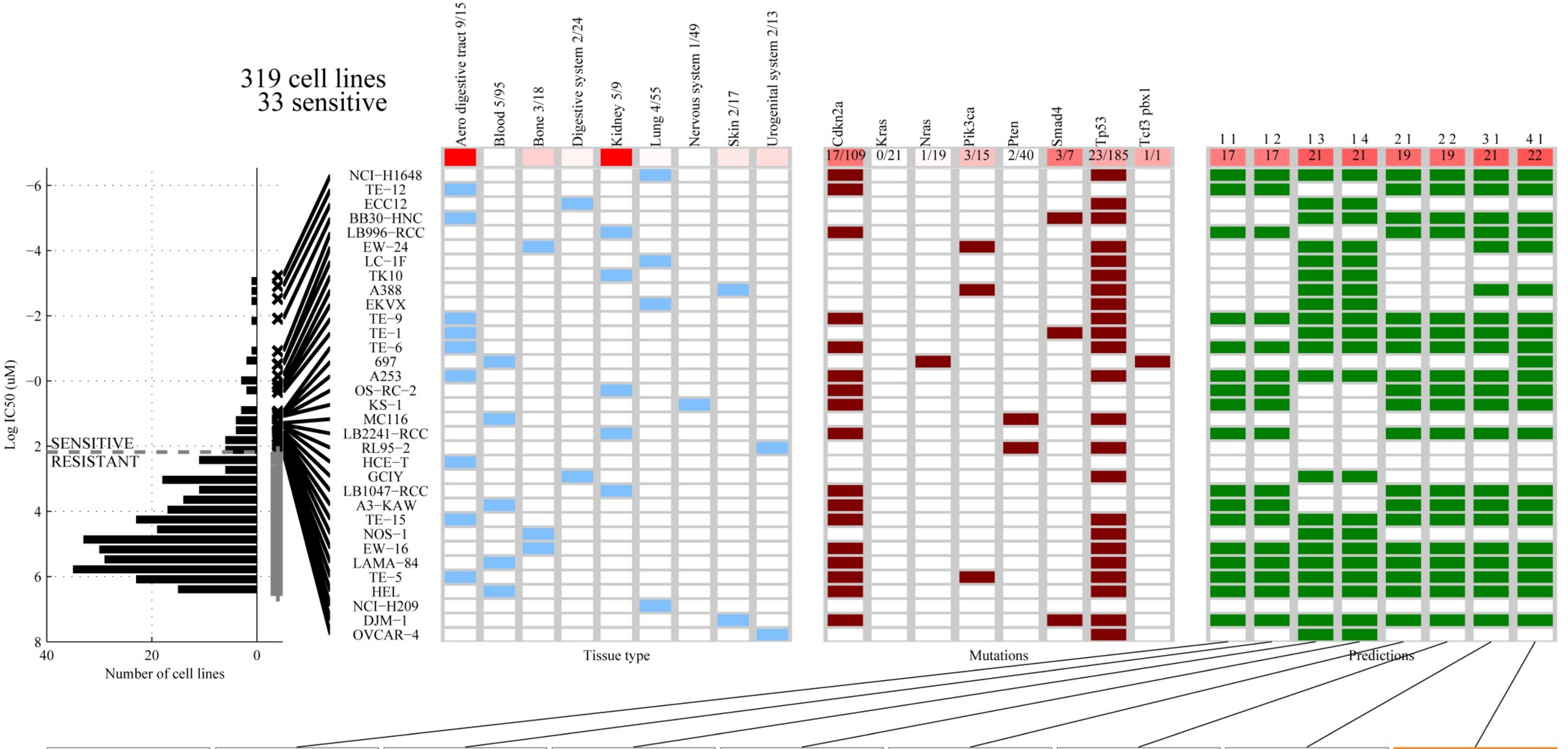
Boxes with the formula and performance statistics for each of the eight logic models. The top part of the box indicates the model complexity (as defined by *K* and *M*), i.e. the type of logic model, from single predictor model to multi-input AND and OR models. The middle part lists the logic formula. &, | and ~ indicate AND, OR and NOT, respectively. The color of a genomic feature indicates its importance with bright red meaning very important and black not important. The lower part gives several performance criteria.

The logic model with the best performance according to cross-validation is highlighted in orange.

	Single predictor	**2-input AND**	**3-input AND**	**4-input AND**	**2-input OR**	**2-by-2**	**3-input OR**	**4-input OR**
Model name K M	1 1 1 1	1 2 1 2	1 3 1 3	1 4 1 4	2 1 2 1	2 2 2 2	3 1 3 1	4 1 4 1
Logic formula	**CDKN2**	**CDKN2&~PTEN**	**~KRAS&~PTEN&** **TP53**	**~KRAS&~NRAS&** **~PTEN& TP53**	**CDKN2 | SMAD4**	**[~KRAS&SMAD4]** **|** **[CDKN2&~PTEN]**	**CDKN2 | PIK3C |** **SMAD4**	**CDKN2 | PIK3C |** **SMAD4 | TCF3**
TP FP FN TN	17 92 16 194	17 75 16 211	21 120 12 166	21 113 12 173	19 95 14 191	19 77 14 209	21 104 12 182	22 104 11 182
Specificity Precision Recall	0.68 0.16 0.52	0.74 0.18 0.52	0.58 0.15 0.64	0.6 0.16 0.64	0.67 0.17 0.58	0.73 0.2 0.58	0.64 0.17 0.64	0.64 0.17 0.67

Complete visualization of the logic models for a single drug inferred by LOBICO

ID:1 Erlotinib –> EGFR



Model name	1 1	1 2	1 3	1 4	2 1	2 2	3 1	4 1
K M	1 1	1 2	1 3	1 4	2 1	2 2	3 1	4 1
Logic formula	**CDKN2**	**CDKN2&~PTEN**	**~KRAS&~PTEN&** **TP53**	**~KRAS&~NRAS&** **~PTEN& TP53**	**CDKN2 | SMAD4**	**[~KRAS&SMAD4]** **|** **[CDKN2&~PTEN]**	**CDKN2 | PIK3C |** **SMAD4**	**CDKN2 | PIK3C |** **SMAD4 | TCF3**
TP FP FN TN	17 92 16 194	17 75 16 211	21 120 12 166	21 113 12 173	19 95 14 191	19 77 14 209	21 104 12 182	22 104 11 182
Specificity Precision Recall	0.68 0.16 0.52	0.74 0.18 0.52	0.58 0.15 0.64	0.6 0.16 0.64	0.67 0.17 0.58	0.73 0.2 0.58	0.64 0.17 0.64	0.64 0.17 0.67