SUPPLEMENTARY DATA

Table S1. Sample number distribution for calibration (cross validation) and test data sets as a function of the 72 Chilean vineyard regions sampled.

		Number of samples	
#	**Vineyard**	**Calibration**	**Test**
1	Agrocauquenes	0	18
2	Agua Santa	24	6
3	Apero	12	0
4	Arboleda	24	0
5	Chomedahue	12	0
6	Clulenco	8	4
7	Colliguay	0	6
8	Conquistadores	12	0
9	Crucesillas/Los conquistadores	12	0
10	Don Melchor	24	12
11	Dos Amalias	18	0
12	El _Boldo	18	0
13	El Arenal	17	6
14	El Aromo	26	4
15	El Boldo	137	71
16	El Delirio	12	0
17	El Estero	15	8
18	El Llano	10	8
19	El Maiten	18	0
20	El Mariscal	12	24
21	El Pillán	0	12
22	El Romero	0	12
23	El Rosario	16	2
24	El Vaticano	10	2
25	Elba Cruz Collligua	12	0
26	Geoagro	4	8
27	Idahue	34	8
28	Julio Ficher	12	0
29	Keule	24	0
30	La Higuera	6	0
31	La Junta	6	0
32	La Puerta	38	14
33	La Rioja	18	0
34	La Trapa	4	8

		Number of samples	
#	**Vineyard**	**Calibration**	**Test**
35	Las Arañas	12	0
36	Las Mercedes	34	8
37	Los Alamos	6	6
38	Los Conquistadores	0	6
39	Lourdes	367	120
40	Manzur	4	2
41	Mariposa	10	2
42	Mariposas	22	2
43	Mariscal	6	6
44	Miraflores	18	0
45	Palo Santo	30	0
46	Panama	8	3
47	Patricio Valdez	10	2
48	Peumo	146	54
49	Pirque	40	6
50	Pocoa	30	0
51	Pozo Merino	8	4
52	Productores	22	5
53	Qda. De Agua	8	4
54	Qta. Maipo	4	2
55	Quebrada de Agua	38	10
56	Quilañanco	0	6
57	Quinta Maipo	18	0
58	R alto Candil	6	6
59	Requinoa	128	51
60	Rucahue	30	6
61	San Adolfo	12	0
62	San Manuel	36	6
63	San Pedro	12	0
64	Santa Delia	23	0
65	Santa Eliana	6	0
66	Santa Olga	12	0
67	Santa Raquel	36	0
68	Santa Sofía	17	0
69	Tocornal	8	16
70	Torretagle	6	0
71	V. Guillaume	42	12
72	V. Alegre	6	6
	sum	1816	584

Table S2. Calibration model fitting parameters for support vector machine (SVM) regression and discriminant analysis including preprocessing, tuning parameters, and cross validation parameters.

	SVM Regression	**SVM Discriminant analysis**
Preprocessing		
X-block	Mean centre GLSW weighting: 0.02	Mean centre GLSW weighting: 0.02
Y-block	Mean centre	N/A
SVM Model Tuning Parameters		
Type	epsilon-SVR	C-SVC
Kernel type	Radial basis function	Radial basis function
Cost	3.1623	100
epsilon	0.01	N/A
gamma	0.031623	0.003162
X-block Compression	PLS (10 Latent Variables)	None
Cross Validation	Class ID (sample replicates)	Class ID (sample replicates)
	Left out data	Left out data
Min	9%	9%
Max	11%	11%
Mean	10%	10%

Table S3. Confusion matrix test results of support vector machine discriminant analysis for the classification of grape samples as a function of an IBMP concentration threshold of 2 ng/kg. *The footnote defines the respective discrimination parameters.

Class	TPR	FPR	TNR	FNR	*n*	Err	P	F1
> 2 ng/kg IBMP	0.9505	0.0000	1.0000	0.0495	101	0.0086	1.0000	0.9746
< 2 ng/kg IBMP	1.0000	0.0495	0.9505	0.0000	483	0.0086	0.9898	0.9949

**Footnote for Table S3*

TPR: proportion of positive cases that were correctly identified (Sensitivity), = TP/(TP+FN)

FPR: proportion of negatives cases that were incorrectly classified as positive, = FP/(FP+TN)

TNR: proportion of negatives cases that were classified correctly (Specificity), = TN/(TN+FP)

FNR: proportion of positive cases that were incorrectly classified as negative, = FN/(FN+TP)

Err: Misclassification error = proportion of samples which were incorrectly classified, = 1-accuracy, = (FP+FN)/(TP+TN+FP+FN)

P: Precision, = TP/(TP+FP)

*F1: F1 Score, = 2*TP/(2*TP+FP+FN)*