COMBINING TWO ANALYTICAL TECHNIQUES WITH CHEMOMETRIC ANALYSIS TO CHARACTERIZE WINE

BY VINEYARD, REGION, AND VINTAGE

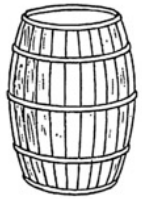
ALEXANDRA CROOK

UNIVERSITY OF NEBRASKA – LINCOLN

University of Nebraska Lincoln

A YEAR IN THE LIFE OF WINE

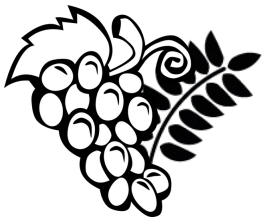
Growth	Mashing	Fermentation	Aging	Bottling
Location, region, soil type, and climate	Growth product is mashed into a paste and cooked Starch → sugar	Combined with yeast Yeast + sugar → ethanol	Alcohol left to age in oak barrels	Bottled and/or blended for sale

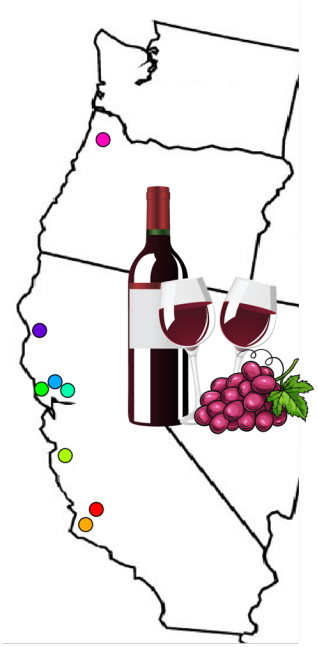


A YEAR IN THE LIFE OF WINE

Growth

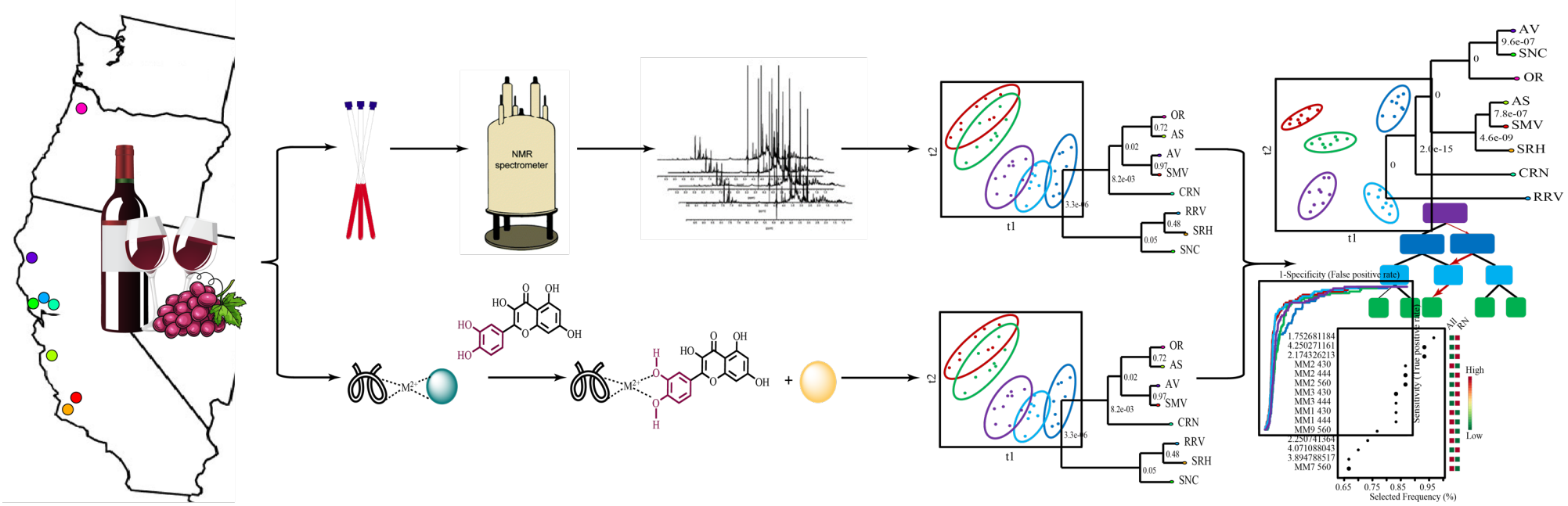
Location, region, soil type, and climate





- Wine grapes (*Vitis vinifera* L. 'Pinot noir' clone 667) from 15 different vineyard sites along the Pacific Coast of the United States
- Grapes were harvested at a similar sugar concentration
  - between 13 August to 15 September 2015 and between 25 August to 21 September 2016
- All wine were developed by the Runnebaum lab at UC Davis

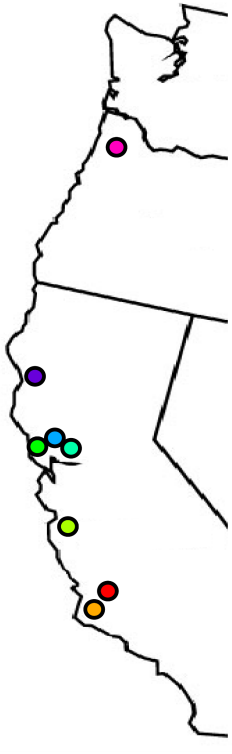
EXPERIMENTAL DESIGN



Wine was developed and produced by the Runnebaum lab at UC Davis and the Differential Sensing data was collected by the Anslyn lab at UT Austin

VINEYARDS, REGIONS, AND VINTAGES

OH MY



VINEYARD	CODE	REGION	2015 n(a)	2015 n(r)	2016 n(a)	2016 n(r)
Nielson	NN	Santa Maria Valley	8	16	8	16
Rice/Cambria	RE	Santa Maria Valley	8		8	
Radian	RN	Santa Rita Hills	8	8	8	8
Panorama 5A	P5A	Arroyo Seco	8	16	8	16
Panorama MSA	PMSA	Arroyo Seco	8		8	
Annapolis	AS	Sonoma Coast	8	16	8	16
Cloud Landing	CL	Sonoma Coast	8		8	
Carneros Hills West	CHW	Sonoma Carneros	6	6	8	8
Ross	RS	Sonoma RRV	8		8	
Bones	BS	Sonoma RRV	7	23	8	24
Bloomfield	BD	Sonoma RRV	8		8	
Boone Ridge	BE	Anderson Valley	7	15	8	16
Maggy Hawk/Falk	MHF	Anderson Valley	8		8	
Gran Moraine	GM	Willimette Valley	8	16	8	16
Zena West	ZW	Willimette Valley	8		8	

EXPERIMENTAL DESIGN - CLASSIFICATION

MULTIVARIATE ANALYSIS

- MVAPACK was utilized for multivariate analysis
  - Raw NMR data was loaded in and processed completely through model validation
  - The DS assay data was uploaded and processed in parallel
  - The two dataset were combined in a multiblock analysis to give equal weight to each technique

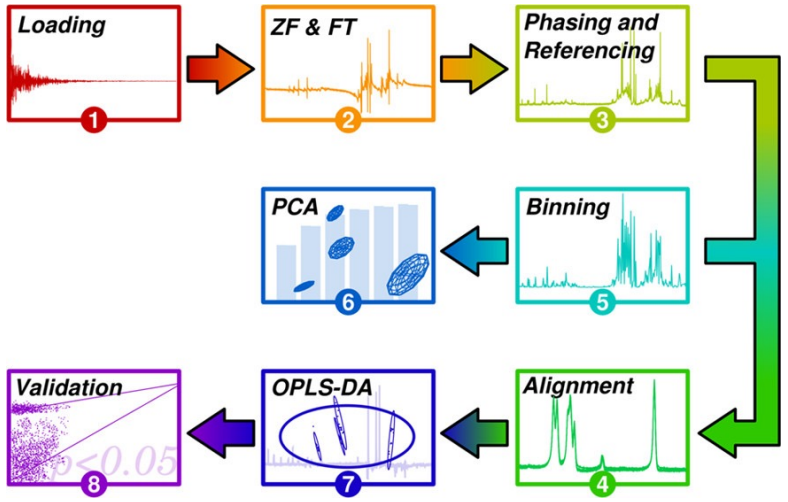
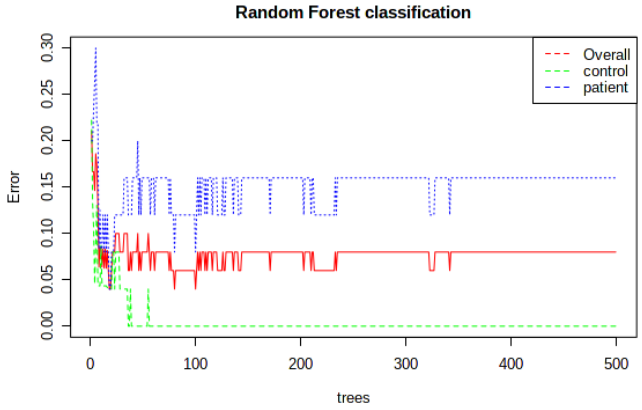


Image from Worley et al. ACS Chem. Biol. 2014, 9, 5, 1138–1144

EXPERIMENTAL DESIGN - CLASSIFICATION

FEATURE ANALYSIS

- MetaboAnalyst is an online chemometrics tools that was utilized for feature analysis
  - Random Forest (RF) was used to test the classification strength of the data based on vineyard, region, and vintage
    - RF was preformed with 500 trees and 7 predictors
    - Out of Bag Error (OOB) used to obtain unbiased estimate of classification error



OOB 0.08

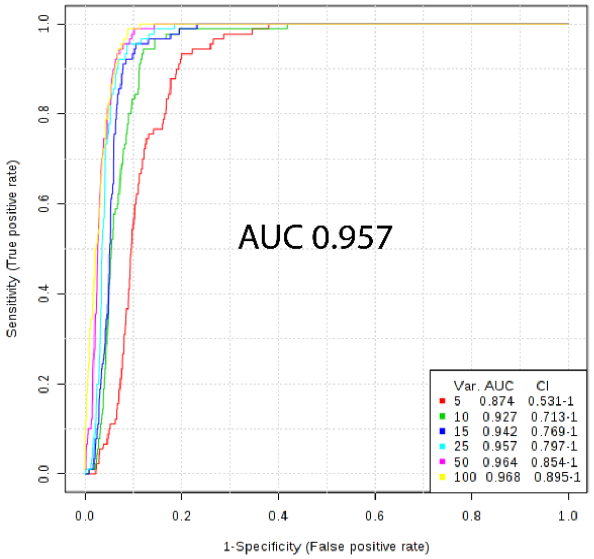
	Control	Patient	Class Error
Control	25	0	0.00
Patient	4	21	0.16

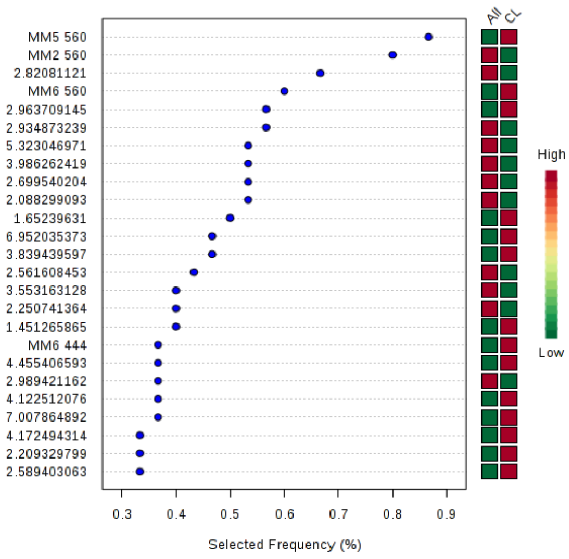
Images provided by Metaboanalyst: Xia, J et al. *Nucleic Acids Res.* **43**, W251–W257 (2015).

EXPERIMENTAL DESIGN - CLASSIFICATION

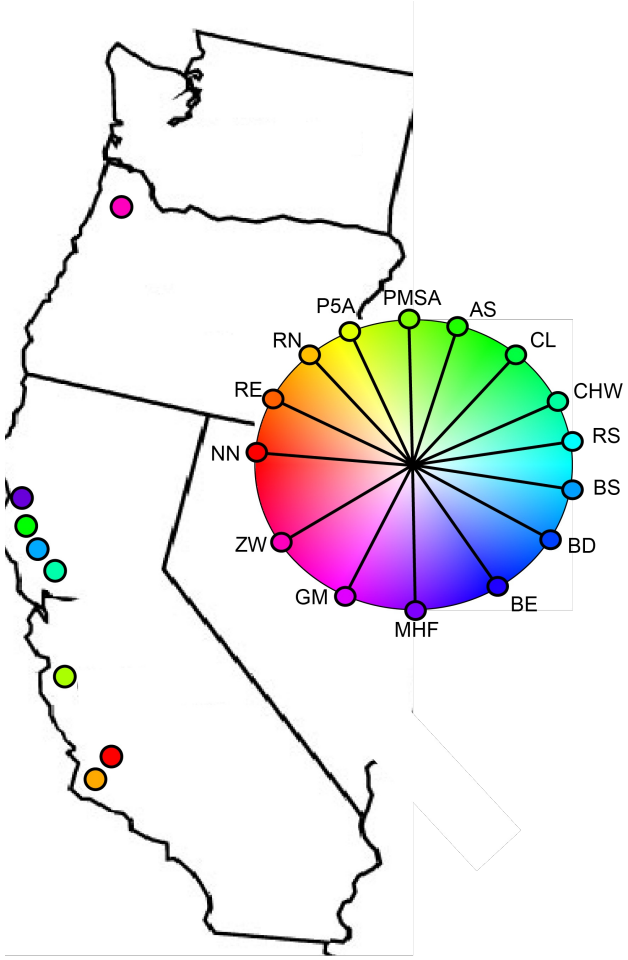
FEATURE ANALYSIS

- MetaboAnalyst is an online chemometrics tools that was utilized for feature analysis
  - ROC Curve Analysis was used to test classification strength and identify important features in each classification





Images provided by Metaboanalyst: Xia, J et al. *Nucleic Acids Res.* **43**, W251–W257 (2015).



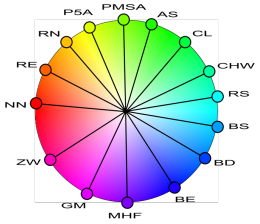
VINEYARD CLASSIFICATION

FIFTEEN VITIS VINIFERA L. 'PINOT NOIR' WINES DERIVED FROM THE SAME SCION CLONE (PINOT NOIR 667)

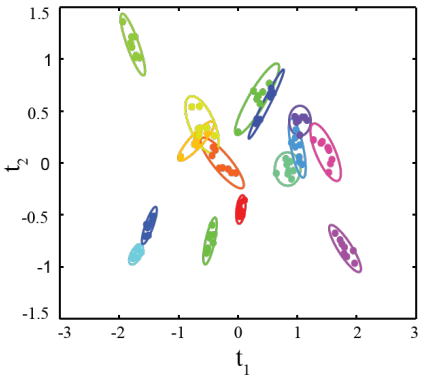
OUR GOAL WAS TO UTILIZE TWO DISTINCT ANALYTICAL TECHNIQUES TO IMPROVE THE CLASSIFICATION BASED ON VINEYARD SPECIFIC CHARACTERISTICS

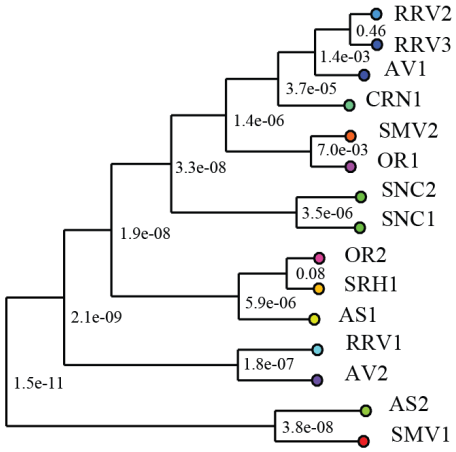
2015 VINEYARD DATA

- PCA analysis was carried out on NMR and Assay data separately
- Combined PCA components used to generate multiblock analysis
- LDA models were generated for models with 4 components

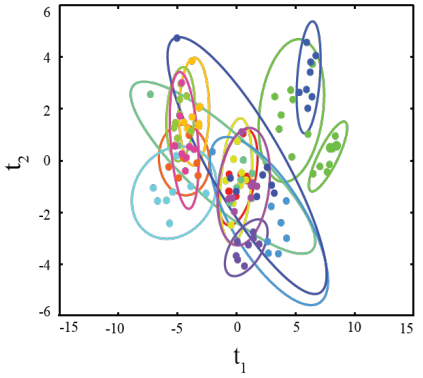


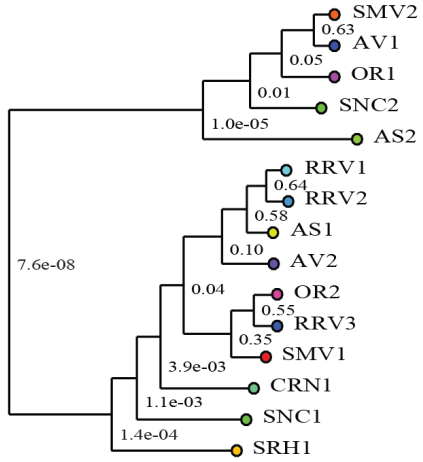
NMR



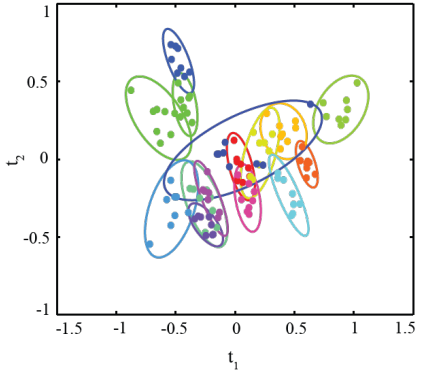


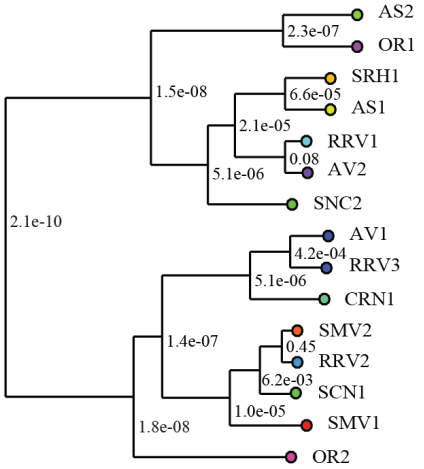
Assay





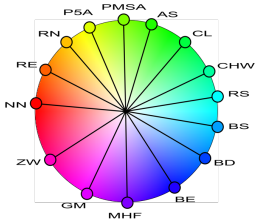
Multiblock



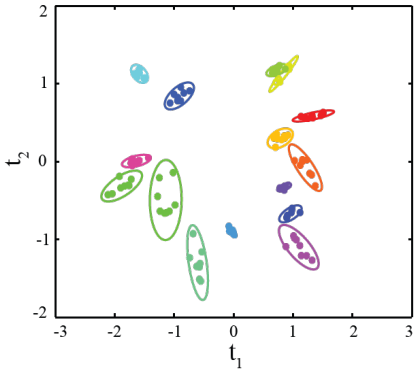


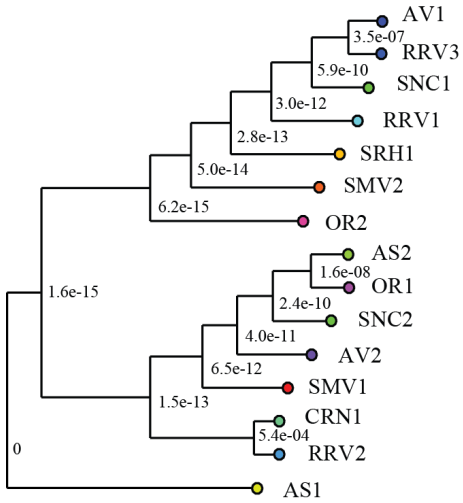
2016 VINEYARD DATA

- PCA analysis was carried out on NMR and Assay data separately
- Combined PCA components used to generate multiblock analysis
- LDA models were generated for models with 4 components

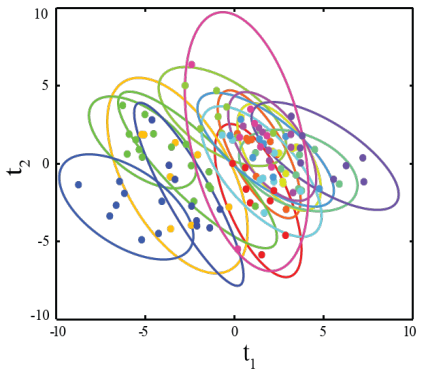


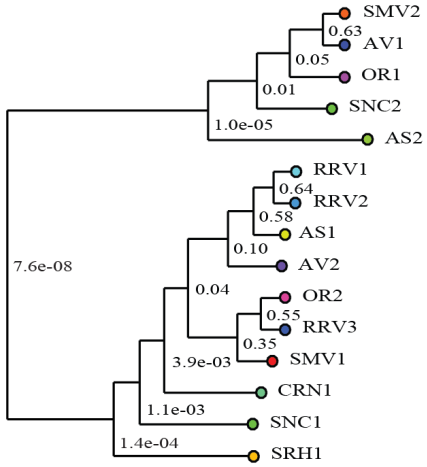
NMR



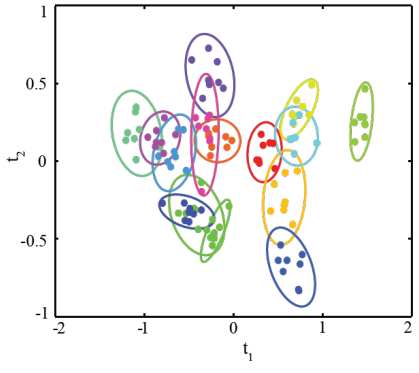


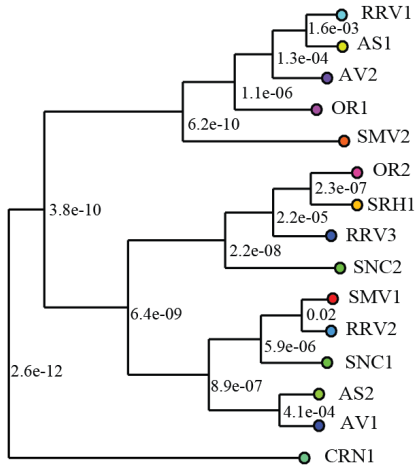
Assay





Multiblock

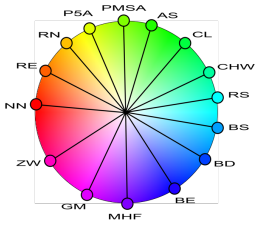




COMBINED NMR+ASSAY DATA IMPROVES VINEYARD CLASSIFICATION

Random Forrest

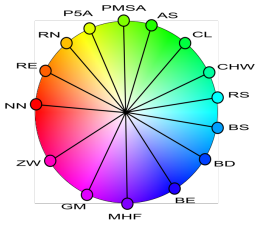
Sample Name		Random Forest Classification					
		NMR	Assay	NMR+Assay	NMR	Assay	NMR+Assay
		2015 Vineyard			2016 Vineyard		
Nielson	NN	0.88	1.00	1.00	1.00	0.75	1.00
Rice/Cambria	RE	1.00	0.88	1.00	0.88	0.88	1.00
Radian	RN	0.75	0.88	0.88	1.00	1.00	1.00
Panorama 5A	P5A	0.63	1.00	0.88	1.00	0.88	1.00
Panorama MSA	PMSA	0.75	0.88	1.00	0.88	0.88	0.88
Annapolis	AS	1.00	0.88	1.00	1.00	1.00	1.00
Cloud Landing	CL	0.75	1.00	1.00	1.00	0.75	1.00
Carneros Hills West	CHW	1.00	1.00	1.00	0.88	0.50	0.88
Ross	RS	1.00	0.75	1.00	1.00	0.75	1.00
Bones	BS	1.00	0.86	1.00	1.00	0.75	1.00
Bloomfield	BD	1.00	1.00	1.00	1.00	1.00	1.00
Boone Ridge	BE	0.86	0.86	0.86	1.00	0.75	1.00
Maggy Hawk/Falk	MHF	0.88	1.00	1.00	1.00	1.00	1.00
Gran Moraine	GM	1.00	1.00	1.00	1.00	1.00	1.00
Zena West	ZW	1.00	0.88	1.00	1.00	0.63	1.00
		0.90 ± 0.12	0.92 ± 0.08	0.97 ± 0.05	0.98 ± 0.05	0.83 ± 0.15	0.98 ± 0.04

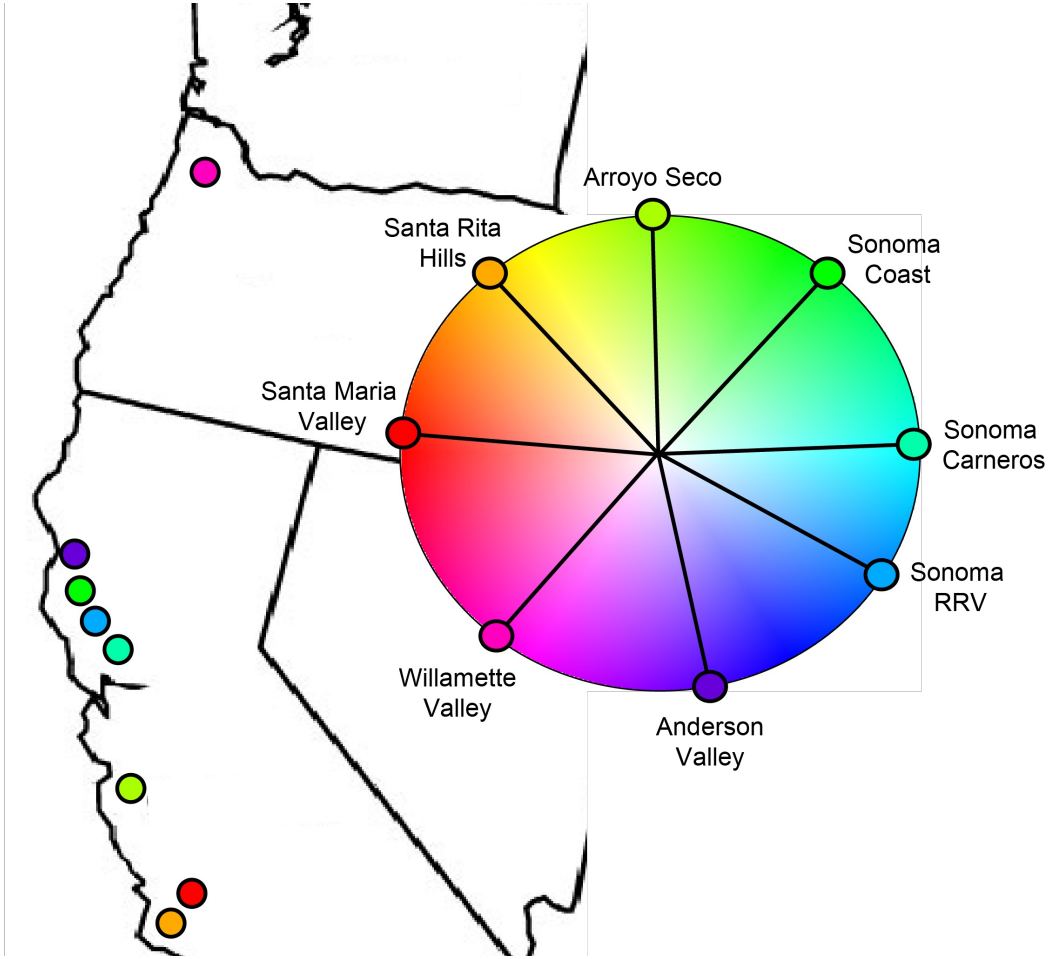


COMBINED NMR+ASSAY DATA IMPROVES VINEYARD CLASSIFICATION

ROC Curve

Sample Name		NMR+Assay ROC					
		AUC	NMR Ratio	Assay Ratio	AUC	NMR Ratio	Assay Ratio
		2015 Vineyard			2016 Vineyard		
Nielson	NN	0.95	0.93 (0.07)	0.07 (0.04)	0.97	0.87 (0.06)	0.13 (0.07)
Rice/Cambria	RE	0.98	0.73 (0.05)	0.27 (0.15)	0.96	0.72 (0.09)	0.28 (0.26)
Radian	RN	0.91	0.4 (0.05)	0.6 (0.56)	0.95	0.64 (0.08)	0.36 (0.33)
Panorama 5A	P5A	0.81	0.88 (0.1)	0.12 (0.11)	0.95	0.47 (0.03)	0.53 (0.3)
Panorama MSA	PMSA	0.98	1 (0.07)	0 (0)	0.95	0.4 (0.03)	0.6 (0.33)
Annapolis	AS	0.98	0.68 (0.08)	0.32 (0.3)	0.97	0.67 (0.05)	0.33 (0.19)
Cloud Landing	CL	0.99	0.2 (0.01)	0.8 (0.44)	0.96	0.84 (0.1)	0.16 (0.15)
Carneros Hills West	CHW	0.92	0.64 (0.08)	0.36 (0.33)	0.99	0.73 (0.05)	0.27 (0.15)
Ross	RS	0.99	0.93 (0.07)	0.07 (0.04)	1.00	1 (0.07)	0 (0)
Bones	BS	0.98	0.52 (0.06)	0.48 (0.44)	0.96	0.88 (0.1)	0.12 (0.11)
Bloomfield	BD	0.99	0.67 (0.05)	0.33 (0.19)	0.99	0.4 (0.02)	0.6 (0.22)
Boone Ridge	BE	0.94	0.92 (0.11)	0.08 (0.07)	1.00	0.87 (0.06)	0.13 (0.07)
Maggy Hawk/Falk	MHF	0.94	0.64 (0.08)	0.36 (0.33)	0.97	0.72 (0.09)	0.28 (0.26)
Gran Moraine	GM	0.99	0.8 (0.06)	0.2 (0.11)	0.99	0.8 (0.04)	0.2 (0.07)
Zena West	ZW	0.95	0.32 (0.04)	0.68 (0.63)	0.99	1 (0.07)	0 (0)



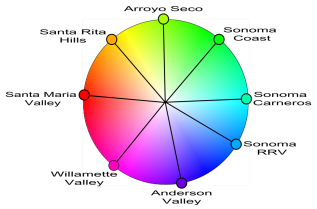


REGION CLASSIFICATION

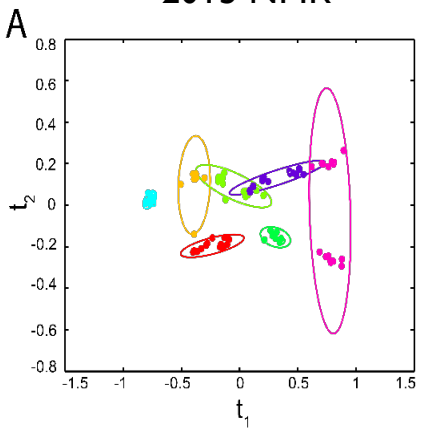
FIFTEEN VITIS VINIFERA L. 'PINOT NOIR' WINES DERIVED FROM THE SAME SCION CLONE (PINOT NOIR 667)

OUR GOAL WAS TO UTILIZE TWO DISTINCT ANALYTICAL TECHNIQUES TO IMPROVE THE CLASSIFICATION BASED ON REGION SPECIFIC CHARACTERISTICS

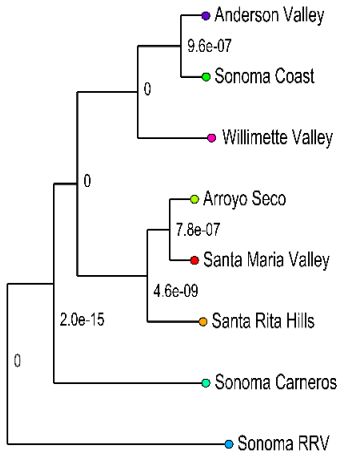
2015 VINEYARD DATA



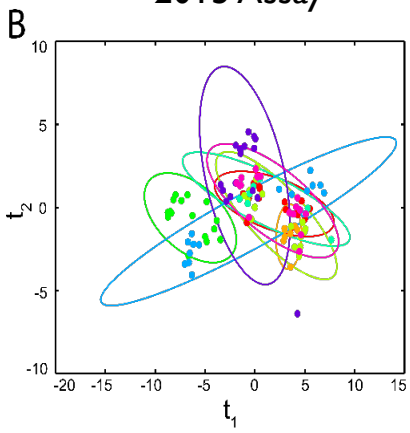
2015 NMR



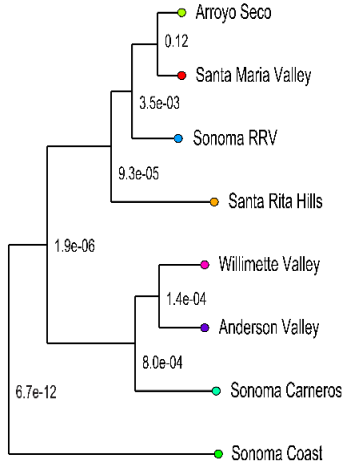
R^2 0.7930 Q^2 0.7398



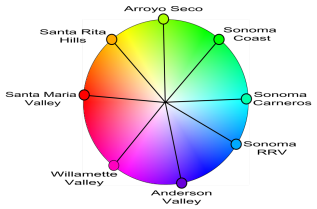
2015 Assay



R^2 0.8796 Q^2 0.7424

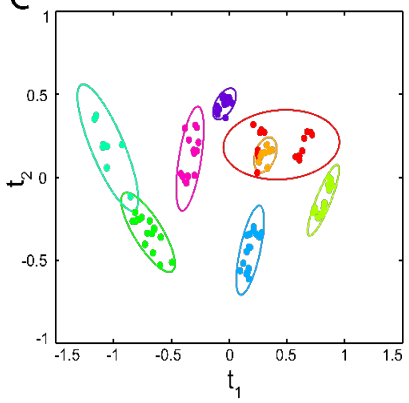


2016 VINEYARD DATA

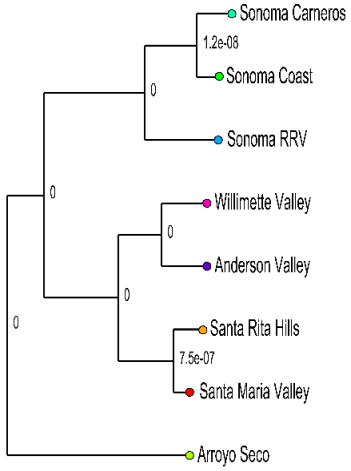


2016 NMR

C

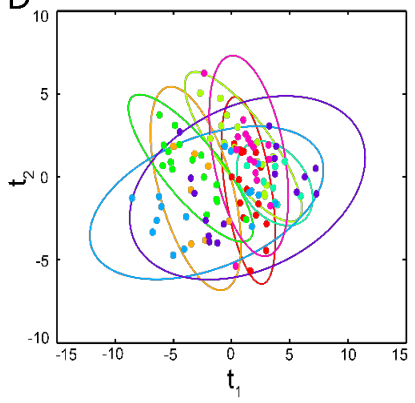


R^2 0.7504 Q^2 0.7014

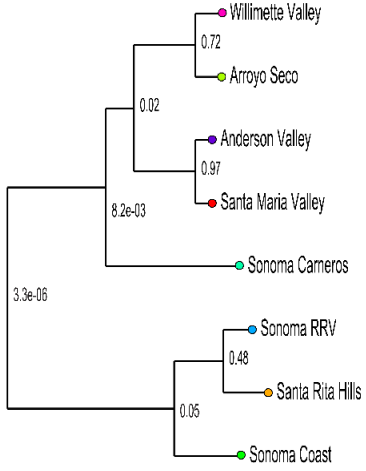


2016 NMR

D

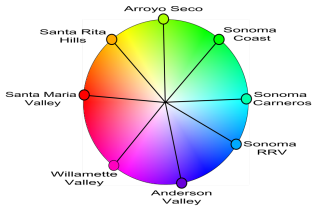


R^2 0.6282 Q^2 0.5209



COMBINED TECHNIQUES PROVIDE IMPROVED CLASSIFICATION BY REGION OF GROWTH

Sample Name		Random Forest Classification					
		NMR	Assay	NMR+Assay	NMR	Assay	NMR+Assay
		2015 Region			2016 Region		
Santa Maria Valley	NN,RE	1.00	0.94	1.00	0.94	1.00	1.00
Santa Rita Hills	RN	0.63	0.63	0.88	1.00	0.75	1.00
Arroyo Seco	P5A,PMSA	0.94	1.00	0.94	1.00	0.94	1.00
Sonoma Coast	AS,CL	0.81	0.81	1.00	1.00	0.88	1.00
Sonoma Carneros	CHW	0.83	0.83	1.00	0.88	0.50	0.88
Sonoma RRV	RS,BS,BD	1.00	0.78	1.00	1.00	0.83	1.00
Anderson Valley	BE,MHF	0.93	0.87	0.93	1.00	0.75	1.00
Willimette Valley	GM,ZW	1.00	0.94	1.00	1.00	0.69	1.00
		0.89±0.12	0.85±0.11	0.97±0.04	0.98±0.04	0.79±0.15	0.98±0.04

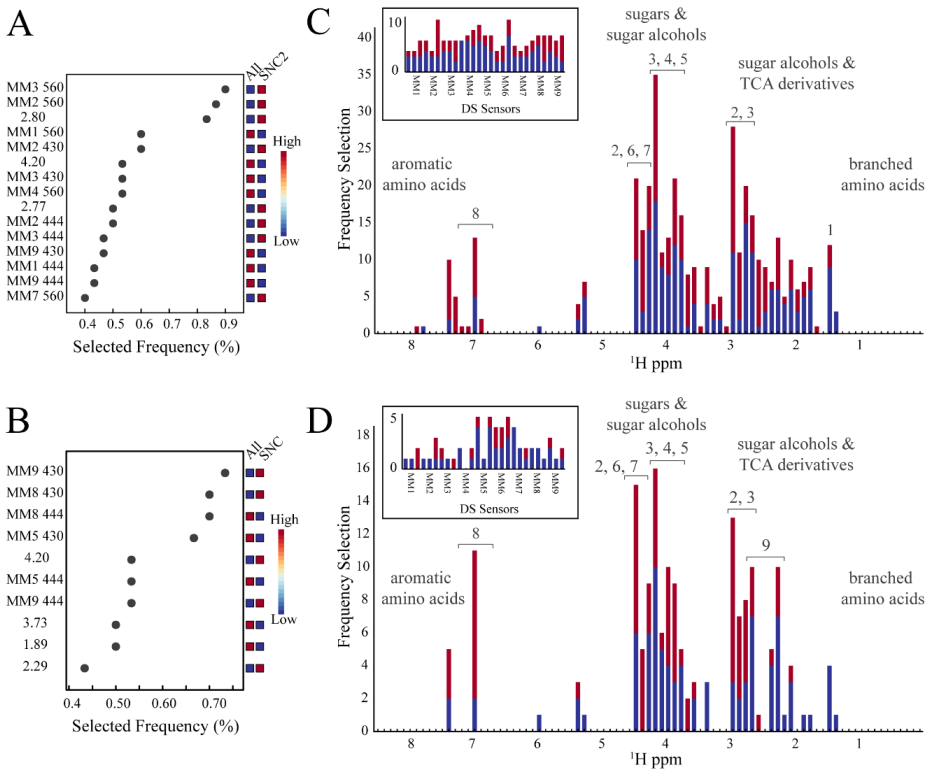


Sample Name		NMR+Assay ROC					
		NMR	Assay	NMR+Assay	NMR	Assay	NMR+Assay
		2015 Region			2016 Region		
Santa Maria Valley	NN,RE	0.97	0.93 (0.07)	0.07 (0.04)	0.99	0.8 (0.04)	0.2 (0.07)
Santa Rita Hills	RN	0.91	0.36 (0.04)	0.64 (0.59)	0.93	0.8 (0.06)	0.2 (0.11)
Arroyo Seco	P5A,PMSA	0.93	0.72 (0.09)	0.28 (0.26)	0.98	0.9 (0.04)	0.1 (0.04)
Sonoma Coast	AS,CL	0.96	0.6 (0.03)	0.4 (0.15)	0.99	0.67 (0.05)	0.33 (0.19)
Sonoma Carneros	CHW	0.87	0.53 (0.04)	0.4 (0.22)	0.99	0.8 (0.06)	0.2 (0.11)
Sonoma RRV	RS,BS,BD	0.99	0.87 (0.06)	0.13 (0.07)	0.98	1 (0.05)	0 (0)
Anderson Valley	BE,MHF	0.96	0.73 (0.05)	0.27 (0.15)	0.99	1 (0.05)	0 (0)
Willimette Valley	GM,ZW	0.99	0.8 (0.04)	0.2 (0.07)	0.98	0.73 (0.05)	0.27 (0.15)

ROC ANALYSIS AND FEATURE SELECTION FREQUENCY

- NMR feature usage from all the ROC curves is plotted using an NMR bin (ppm) size of 0.1 ppm for (C) vineyard and (D) region analysis.
- 2015 data are colored blue, and the 2016 data is colored red.
- A plot of the DS array feature (MM1 to MM9) usage from the same ROC curve analyses are displayed as an insert

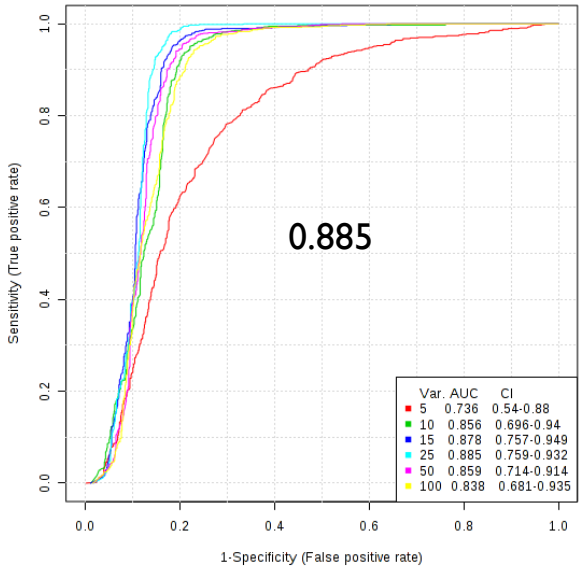
- Putative metabolite IDs correspond to 1, isobutanol; 2, malic acid; 3, phenethyl alcohol; 4, mannitol; 5, fructose; 6, ethyl acetate; 7, ethyl lactate; 8, tyrosine; and 9, citric acid.

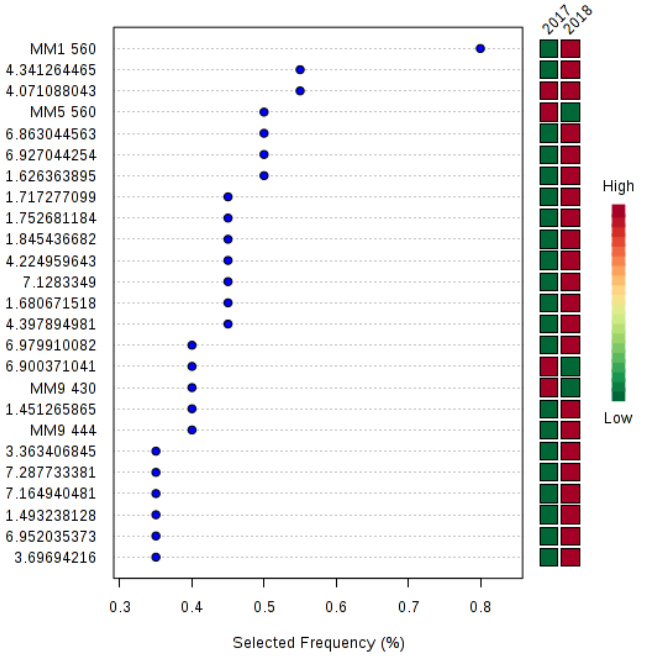


VINTAGE YEAR COMPARISONS

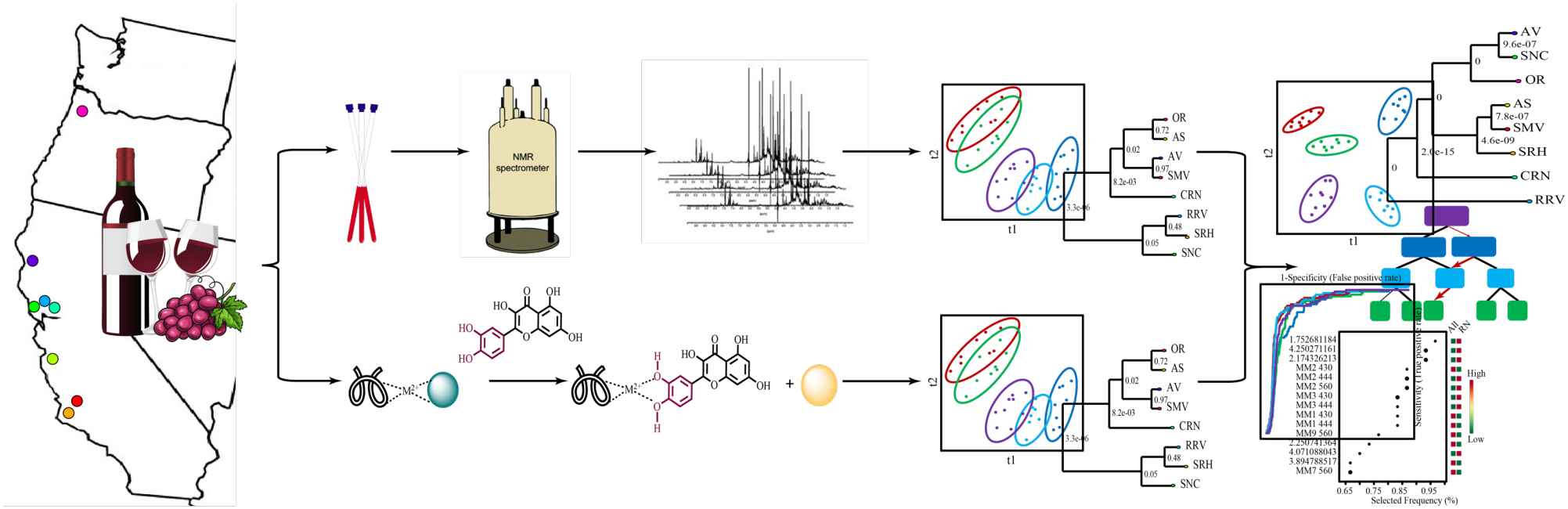
ROC Curve Analysis demonstrated 88.5% AUC with a model containing 25 features

16% Assay Contribution
84% NMR Contribution





COMBINATION OF NMR AND COLORIMETRIC SENSOR FINGERPRINTING IMPROVES WINE CLASSIFICATION



ACKNOWLEDGMENTS

A special thank you to our collaborates Dr. Eric Anslyn and Dr. Diana Zamora-Olivares at the University of Texas Austin

And the students in the Freshmen Research Initiative at the University of Texas at Austin

Special thank you to Dr. Ron Runnebaum at UC Davis as well