

TRGN 599: Applied Data Science and Bioinformatics

UNIT VI. Enrichment Analysis, Linear Regression

Week 14 - Lecture 1

Enrique I. Velazquez Villarreal, M.D., Ph.D., M.P.H., M.S. | Assistant Professor

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Leader of the USC Bioinformatics Core – *USC CaRE2 Health Equity Center*

David W. Craig, Ph.D. | Professor and Vice Chair

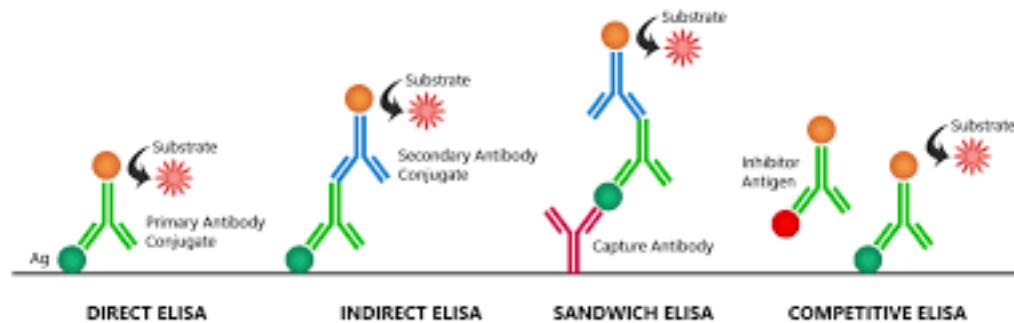
Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Co-Director, Institute of Translational Genomics

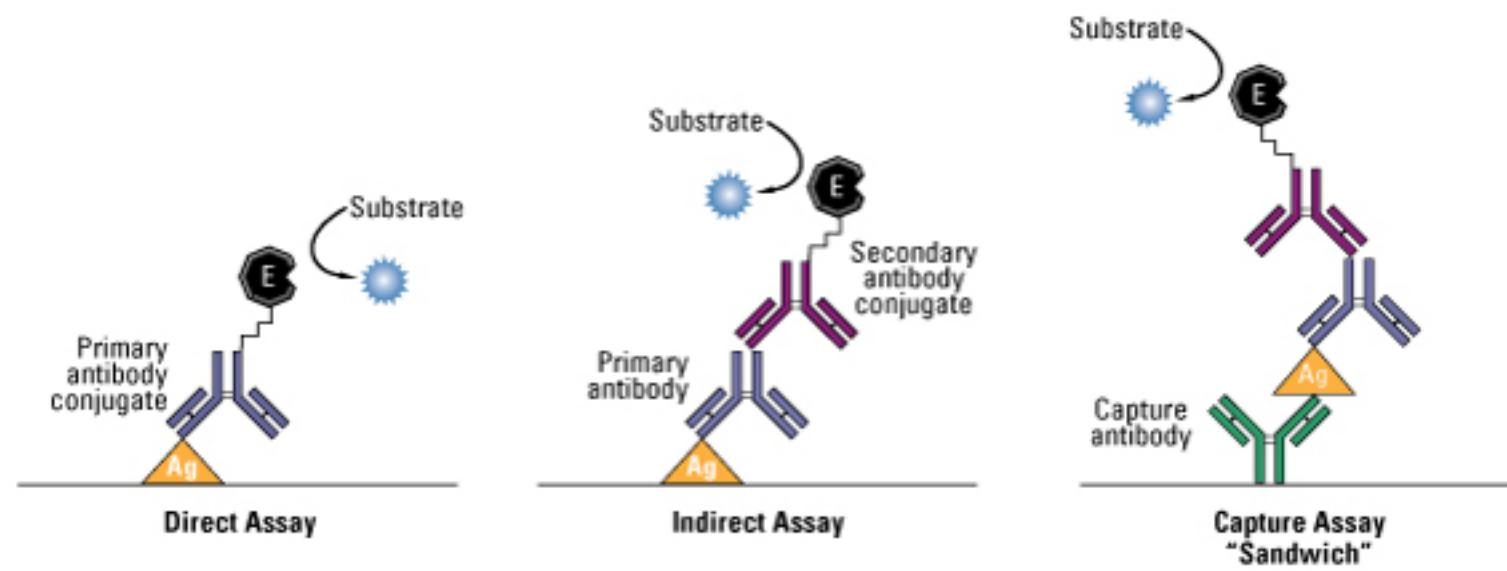
Topics

- Concentration Calculation with ELISA



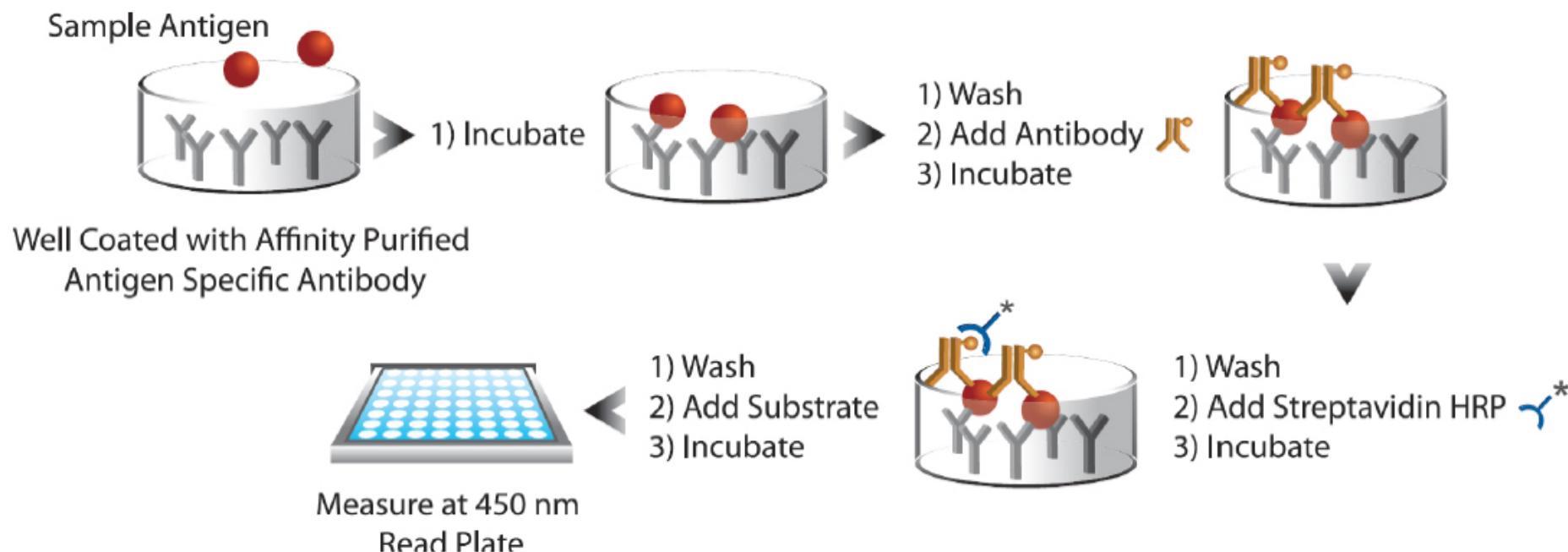
Background

- Enzyme-linked immunosorbent assay (ELISA) is a quantification method for measuring concentration of any kind of molecular compound from biological liquids such as blood, serum, or cell culture supernatants.
- This method is used in molecular biology for a long time, and medium-throughput instruments are available to produce measurement data for dozens of parallel samples.
- Data analysis aspects of ELISA are far from being trivial mostly because of the complicated mathematics of translating measured raw data to concentrations.



Background

- Enzyme-linked immunosorbent assay (ELISA) is a quantification method for measuring concentration of any kind of molecular compound from biological liquids such as blood, serum, or cell culture supernatants.
- This method is used in molecular biology for a long time, and medium-throughput instruments are available to produce measurement data for dozens of parallel samples.
- Data analysis aspects of ELISA are far from being trivial mostly because of the complicated mathematics of translating measured raw data to concentrations.



Background

- Measuring concentrations of peptides, proteins, or other analytes in different biological systems is a fundamental tool of molecular biology.
- In molecular biology, immunoassays are one of the most fundamental tools for measuring the amount of a molecular component specifically.
- The most widely used immunoassay is named as enzyme-linked immunosorbent assay or ELISA.

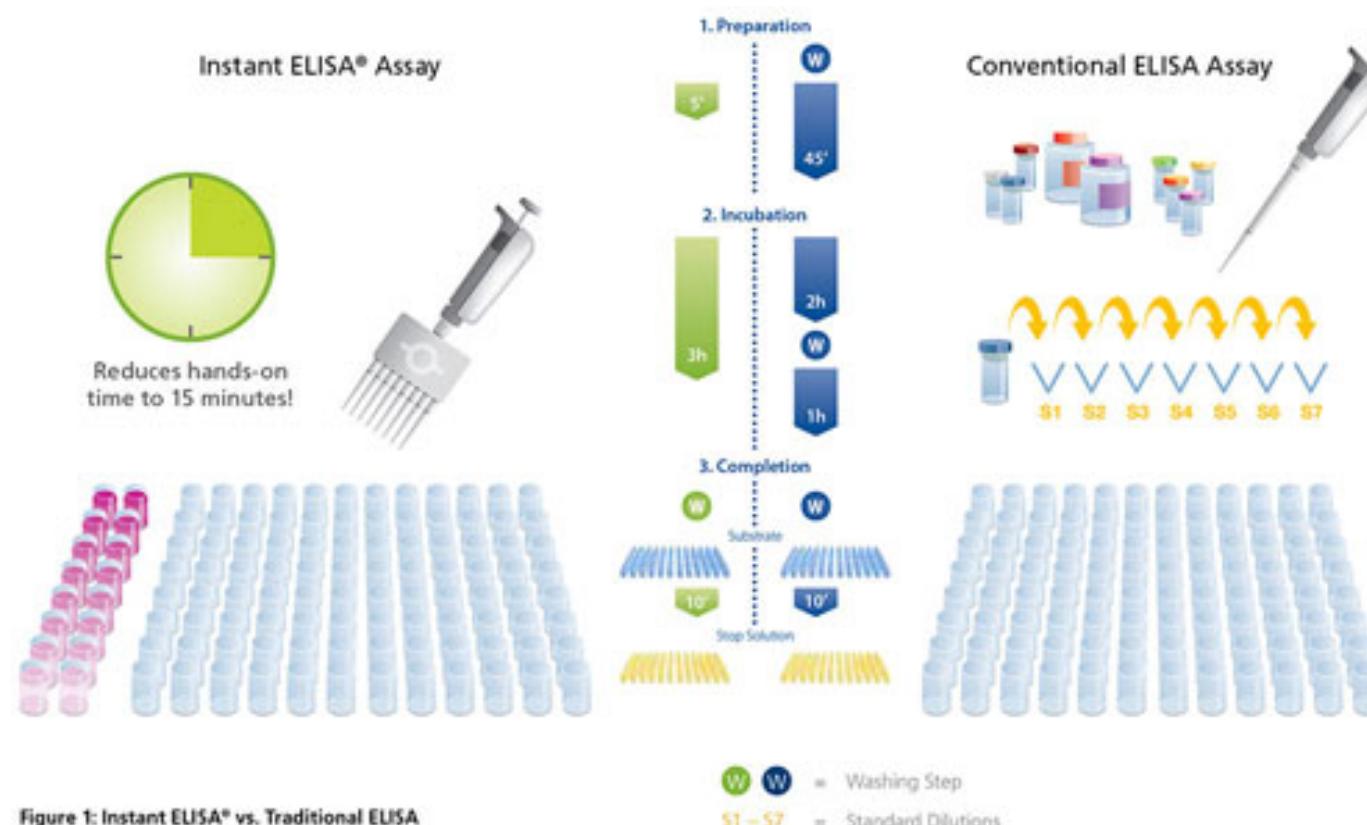
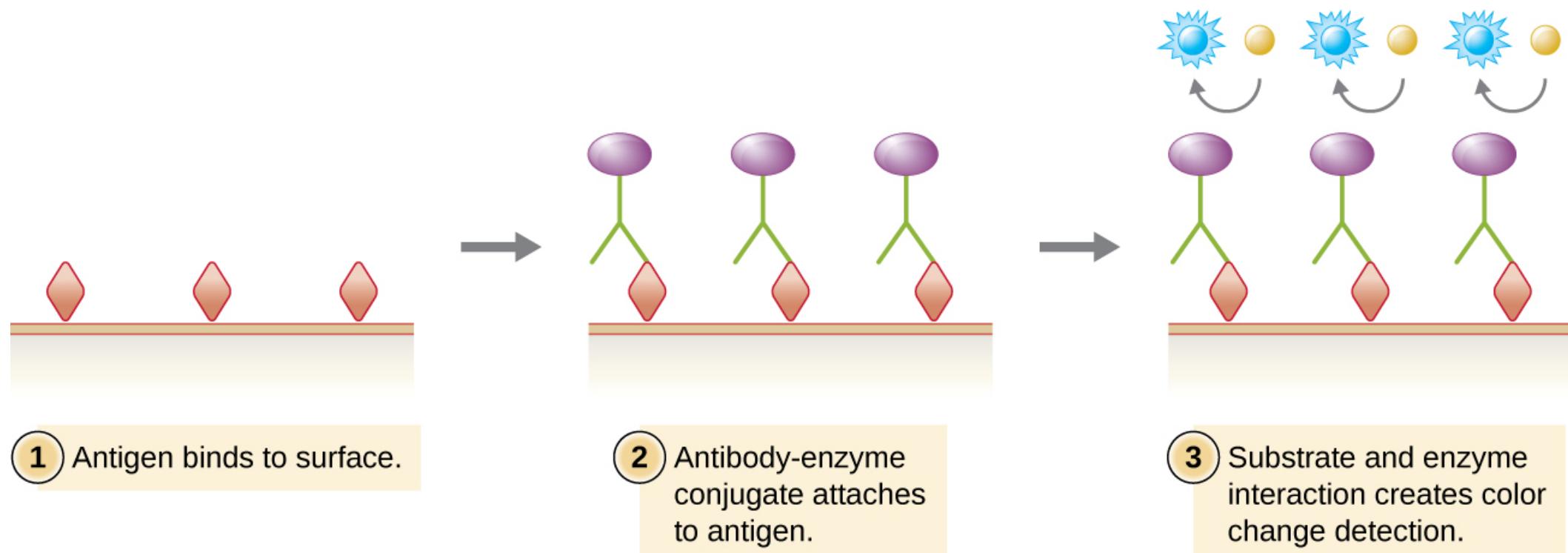


Figure 1: Instant ELISA® vs. Traditional ELISA

Accessing ELISA data

- ELISA experiments are traditionally performed on 96-well microtiter plates (microplates)
- Modern readers have software bundles capable of exporting data in standard file formats, most often in Excel files.
- Nowadays, the transfer of data is a matter of exporting Excel or CSV files, or, in an automated setting, the reader is integrated into a standard laboratory information management system (LISM), and the measured data are uploaded to a relational database for further analysis.



Concentration calculation

- The immediate goal of ELISA data analysis is to calculate the concentration of the target molecule by using known standards.
- This procedure involves curve fitting and interpolation, all established and well-supported operations in R.
- The file can be accessed by using functions from the `readxl` package.

```
8 - ````{R}
9
10 setwd("/Users/enriquevelazquez/Documents/R_working_directory")
11
12 library(readxl)
13
14 xls.file <- "090714_data.xlsx"
15 excel_sheets(xls.file)
16 ````
```

Concentration calculation

- The file contain three sheets
- Two of them, Ch450 and Ch570
 - Contain color absorbance data measured at 450 and 570 nm wavelengths.
- Data is arranged in a matrix format
 - Rows and columns of the microtiter plate.
- Individual reads can be separated into dedicated variables

```
18 ````{R}
19 ch450 <- read_excel(xls.file,sheet=1)
20 ch570 <- read_excel(xls.file,sheet=2)
21 samples <- read_excel(xls.file,sheet=3)
22 ch450
23 ````
```

...1 <chr>	1 <dbl>	2 <dbl>	3 <dbl>	4 <dbl>	5 <dbl>	6 <dbl>	7 <dbl>	8 <dbl>	9 <dbl>
A	0.045	0.066	0.047	3.965	3.965	0.046	0.050	0.047	0.050
B	0.287	0.072	0.069	0.294	0.330	2.151	2.204	0.133	0.139
C	0.294	4.000	2.921	4.000	4.000	1.323	1.364	4.000	4.000
D	0.091	4.001	4.001	4.001	4.001	0.752	0.736	0.136	0.138
E	0.091	1.069	0.986	0.274	0.287	0.430	0.416	3.304	3.304
F	0.339	3.985	3.985	3.985	3.719	0.241	0.221	0.060	0.064
G	0.401	3.992	3.992	3.992	3.992	0.139	0.151	1.650	1.522
H	0.611	0.522	3.958	3.958	3.958	3.958	0.468	0.362	3.692

8 rows | 1-10 of 13 columns

Concentration calculation

- The first column in each data frame are the row identifiers of the plate
 - These are not needed for data analysis.
- Let us delete them by assigning NULL to their column.

```
25 ````{R}
26 ch450[,1] <- NULL
27 ch570[,1] <- NULL
28 samples[,1] <- NULL
29 ch450
30 ````
```

1	2	3	4	5	6	7	8	9	10
<dbl>									
0.045	0.066	0.047	3.965	3.965	0.046	0.050	0.047	0.050	0.052
0.287	0.072	0.069	0.294	0.330	2.151	2.204	0.133	0.139	0.295
0.294	4.000	2.921	4.000	4.000	1.323	1.364	4.000	4.000	4.000
0.091	4.001	4.001	4.001	4.001	0.752	0.736	0.136	0.138	0.382
0.091	1.069	0.986	0.274	0.287	0.430	0.416	3.304	3.304	3.304
0.339	3.985	3.985	3.985	3.719	0.241	0.221	0.060	0.064	0.066
0.401	3.992	3.992	3.992	3.992	0.139	0.151	1.650	1.522	0.072
0.611	0.522	3.958	3.958	3.958	3.958	0.468	0.362	3.692	3.449

8 rows | 1-10 of 12 columns

Concentration calculation

- The created 3 variables will be together assigned as columns of a new dataframe.
- Since matrices in the Excel sheet are interpreted as lists of columns during importing, One should use the unlist() function to convert them to vectors.

```
31
32  ````{R}
33  elisa.data <- data.frame(ch450=unlist(ch450),ch570=unlist(ch570),sample=unlist(samples))
34  head(elisa.data)
35  ````
```

	ch450	ch570	sample
11	0.045	0.040	solvent
12	0.287	0.043	WT1 0h
13	0.294	0.041	WT1 0h
14	0.091	0.041	K03 0h
15	0.091	0.044	K03 0h
16	0.339	0.042	K04 0h

Concentration calculation

- Connecting data from the three sheets of the original data Excel file to a single data frame offers a good opportunity for data handling.
- The individual wells of the microtiter plate are represented by rows.
- We can have a quick inspection on the spread measured values by applying general R functions.

```
36
37 ```{R}
38 summary(elisa.data$ch450)
39 ```

      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
0.045   0.091   0.423   1.602   3.958   4.001

40
41 ```{R}
42 summary(elisa.data$ch570)
43 ```

      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
0.03800 0.04200 0.04600 0.04943 0.05525 0.10600
```

Concentration calculation

- The 570 nm channel measured hardly any difference among the individual wells, while the 450 nm channel registered much more variability.
- The well-to-well variance can be nicely visualized with images mimicking microtiter plates for data representation.

```
45 - ````{R}
46  image(1:12,1:8,t(matrix(elisa.data$ch570,nrow=8)),xlab="",ylab="",ylim=c(8.5,0.5),main="Channel 570")
47  text(rep(1:12,each=8),rep(1:8,12),labels=matrix(elisa.data$sample,nrow=8),pos=3)
48  text(rep(1:12,each=8),rep(1:8,12),labels=matrix(elisa.data$ch570,nrow=8),pos=1)
49  ````
```

Concentration calculation

Channel 570

	solvent	solvent	solvent	KO2 48h	KO2 48h	solvent	solvent	solvent	solvent	solvent	solvent	solvent
2	0.04	0.04	0.041	0.057	0.059	0.039	0.042	0.04	0.039	0.041	0.043	0.044
	WT1 0h	WT2 4h	WT2 4h	WT4 4h	WT4 4h	Std 200 pg/ml	Std 200 pg/ml	WT5 4h	WT5 4h	KO5 4h	KO5 4h	medium
	0.043	0.046	0.044	0.045	0.047	0.051	0.061	0.049	0.044	0.051	0.049	0.053
	WT1 0h	WT2 24h	WT2 24h	WT4 24h	WT4 24h	Std 100 pg/ml	Std 100 pg/ml	WT5 48h 1/10	WT5 48h 1/10	KO5 48h 1/10	KO5 48h 1/10	medium
	0.041	0.047	0.046	0.057	0.056	0.042	0.043	0.055	0.054	0.058	0.058	0.044
4	KO3 0h	WT2 48h 1/10	WT2 48h 1/10	WT4 48h 1/10	WT4 48h 1/11	Std 50 pg/ml	Std 50 pg/ml	WT6 4h	WT6 4h	KO6 4h	KO6 4h	medium
	0.041	0.061	0.059	0.061	0.061	0.044	0.044	0.042	0.042	0.045	0.046	0.106
	KO3 0h	WT1 4h	WT1 4h	KO2 4h	KO2 4h	Std 25 pg/ml	Std 25 pg/ml	WT6 48h 1/10	WT6 48h 1/10	KO6 48h 1/10	KO6 48h 1/10	medium
	0.044	0.05	0.048	0.045	0.047	0.044	0.046	0.06	0.058	0.068	0.065	0.051
6	KO4 0h	WT1 24h	WT1 24h	KO2 24h	KO2 24h	Std 12.5 pg/ml	Std 12.5 pg/ml	WT7 4h	WT7 4h	KO3 48h	KO3 48h	medium
	0.042	0.063	0.056	0.06	0.058	0.04	0.041	0.042	0.041	0.044	0.042	0.048
	KO4 0h	WT1 48h 1/10	WT1 48h 1/10	KO2 48h 1/10	KO2 48h 1/10	Std 6.25 pg/ml	Std 6.25 pg/ml	WT7 48h 1/10	WT7 48h 1/10	KO4 48h	KO4 48h	medium
	0.041	0.063	0.061	0.055	0.055	0.04	0.042	0.046	0.045	0.044	0.043	0.051
8	KO3 4h	KO3 4h	KO3 24h	KO3 24h	KO3 48h 1/10	KO3 48h 1/10	KO4 4h	KO4 4h	KO4 24h	KO4 24h	KO4 48h 1/10	KO4 48h 1/10
	0.038	0.04	0.051	0.051	0.056	0.053	0.038	0.038	0.048	0.05	0.054	0.078
	2	4	6	8	10	12						

Concentration calculation

- On the images, the yellow and white cells represent wells with the highest values.
- Color absorbances measured at 570 nm are very even as expected.
- It is also excellent that no systematic errors, such as too low values from border wells or row-by-row or column-by-column growing values, are visible.
- These images are very handy for spotting such errors usually originating from pipetting or instrument problems.

```
51 - ````{R}
52  image(1:12,1:8,t(matrix(elisa.data$ch450,nrow=8)),xlab="",ylab="",ylim=c(8.5,0.5),main="Channel 450")
53  text(rep(1:12,each=8),rep(1:8,12),labels=matrix(elisa.data$sample,nrow=8),pos=3)
54  text(rep(1:12,each=8),rep(1:8,12),labels=matrix(elisa.data$ch450,nrow=8),pos=1)
55  ````
```

- The coloring of this image is adjusted with the col parameter of image() so that the more yellow cells on the image correspond to the more yellow wells on the microplate.
- Some researchers find it easier to interpret the results this way, connecting numbers and the physical plate visually

Concentration calculation with a standard curve

Channel 450

	solvent	solvent	solvent	KO2 48h	KO2 48h	solvent	solvent	solvent	solvent	solvent	solvent	solvent	
	0.045	0.066	0.047	3.965	3.965	0.046	0.05	0.047	0.05	0.052	0.07	0.088	
2	WT1 0h 0.287	WT2 4h 0.072	WT2 4h 0.069	WT4 4h 0.294	WT4 4h 0.33	Std 200 pg/ml 2.151	Std 200 pg/ml 2.204	WT5 4h 0.133	WT5 4h 0.139	KO5 4h 0.295	KO5 4h 0.366	medium 0.08	
	WT1 0h 0.294	WT2 24h 4	WT2 24h 2.921	WT4 24h 4	WT4 24h 4	Std 100 pg/ml 1.323	Std 100 pg/ml 1.364	WT5 48h 1/10 4	WT5 48h 1/10 4	KO5 48h 1/10 4	KO5 48h 1/10 4	medium 0.072	
4	KO3 0h 0.091	WT2 48h 1/10 4.001	WT2 48h 1/10 4.001	WT4 48h 1/10 4.001	WT4 48h 1/11 4.001	Std 50 pg/ml 0.752	Std 50 pg/ml 0.736	WT6 4h 0.136	WT6 4h 0.138	KO6 4h 0.382	KO6 4h 0.366	medium 0.143	
	KO3 0h 0.091	WT1 4h 1.069	WT1 4h 0.986	KO2 4h 0.274	KO2 4h 0.287	Std 25 pg/ml 0.43	Std 25 pg/ml 0.416	WT6 48h 1/10 3.304	WT6 48h 1/10 3.304	KO6 48h 1/10 3.304	KO6 48h 1/10 3.266	medium 0.083	
6	KO4 0h 0.339	WT1 24h 3.985	WT1 24h 3.985	KO2 24h 3.985	KO2 24h 3.719	Std 12.5 pg/ml 0.241	Std 12.5 pg/ml 0.221	WT7 4h 0.06	WT7 4h 0.064	KO3 48h 0.066	KO3 48h 0.072	medium 0.078	
	KO4 0h 0.401	WT1 48h 1/10 3.992	WT1 48h 1/10 3.992	KO2 48h 1/10 3.992	KO2 48h 1/10 3.992	Std 6.25 pg/ml 0.139	Std 6.25 pg/ml 0.151	WT7 48h 1/10 1.65	WT7 48h 1/10 1.522	KO4 48h 0.072	KO4 48h 0.083	medium 0.081	
8	KO3 4h 0.611	KO3 4h 0.522	KO3 24h 3.958	KO3 24h 3.958	KO3 48h 1/10 3.958	KO3 48h 1/10 3.958	KO4 4h 0.468	KO4 4h 0.362	KO4 24h 3.692	KO4 24h 3.449	KO4 48h 1/10 3.958	KO4 48h 1/10 3.595	
	2	4	6	8	10	12							

Concentration calculation

- Thank you!