U-Statistics for Multiple Censored Outcomes with Varying Frequency, Severity, Attribution

KNUT M. WITTKOWSKI

The Rockefeller University Hospital
Center for Clinical and Translational Research
New York, NY 10021

November 5, 2013

Acknowledgements

- Tingting Song
- Vikas Sonakya
- Martin P. Seybold
- Benedetta Bigio
- Carla Nasca
- Manuel Ascano

- Jose Morales (Columbia)
- Martina Durner (MSSM)
- James G. Krueger (Rockefeller)
- Marion Schneider (Ulm)
- Frederik Shik (Yale)
- Mona Katharina Tonn (RheinAhr)
- Gabrielle Gold-von Simson (NYU)

Doctors Fear Losing Leukemia Drug Deemed Risky

GRADY D, POLLAK A (2013) *NYTimes* Nov 1:A3

An FDA spokeswoman said that studies of the drug to date involved 530 patients of which

2.6 percent (14) had

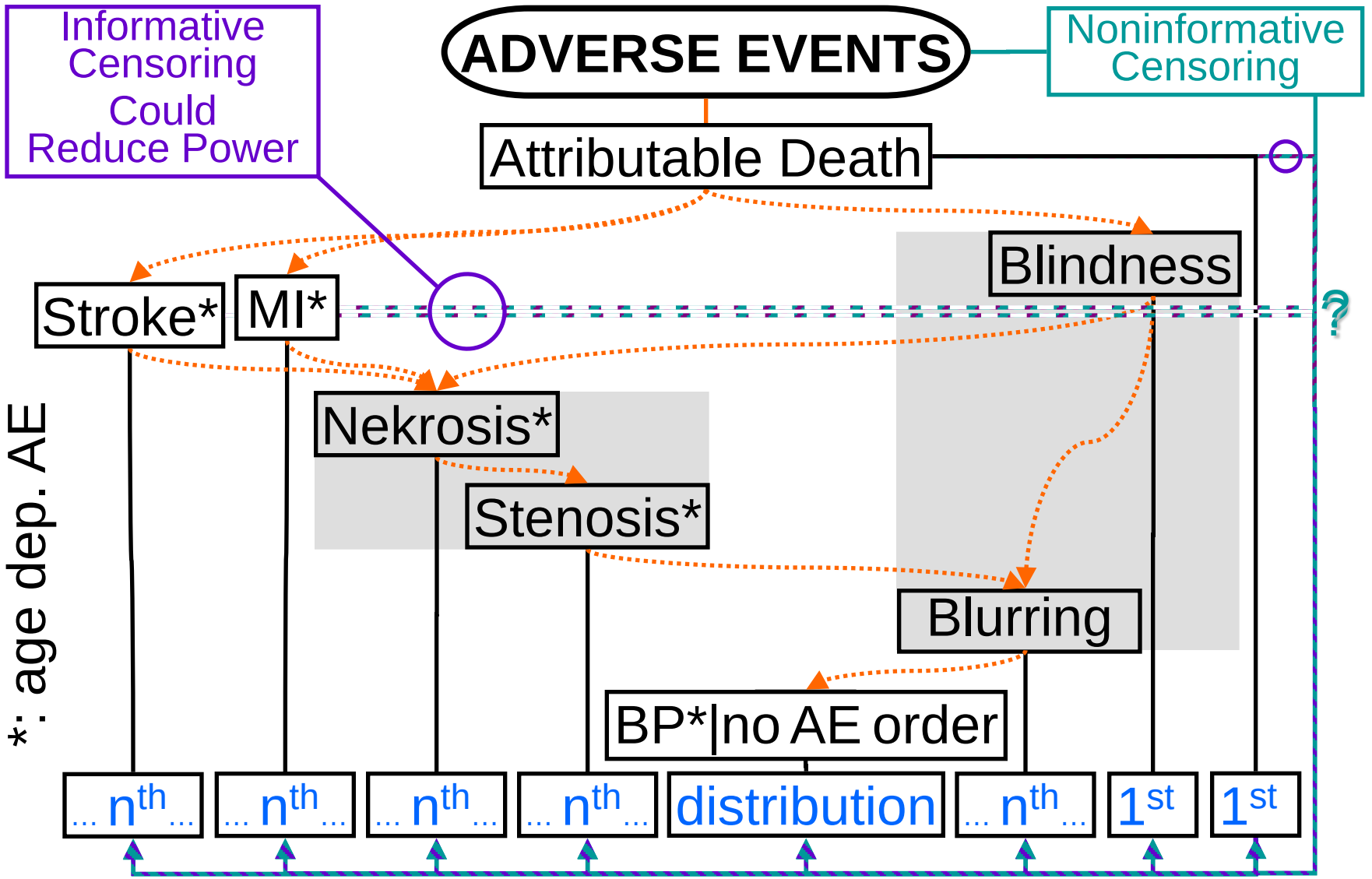
- *died from cardiovascular problems.*

24 percent taking Iclusig for a median of 1.3 years, and 48 percent studied for a median of 2.7 years, had suffered serious adverse vascular events, starting at 2 weeks of treatment, including

- *heart attacks,*
- *strokes,*
- [*blurred vision /*] *blindness*, and
- *lack of blood flow to* [*/ tissue death in*] *the extremities*;

[
- *high blood pressure occurred in 67 percent]*

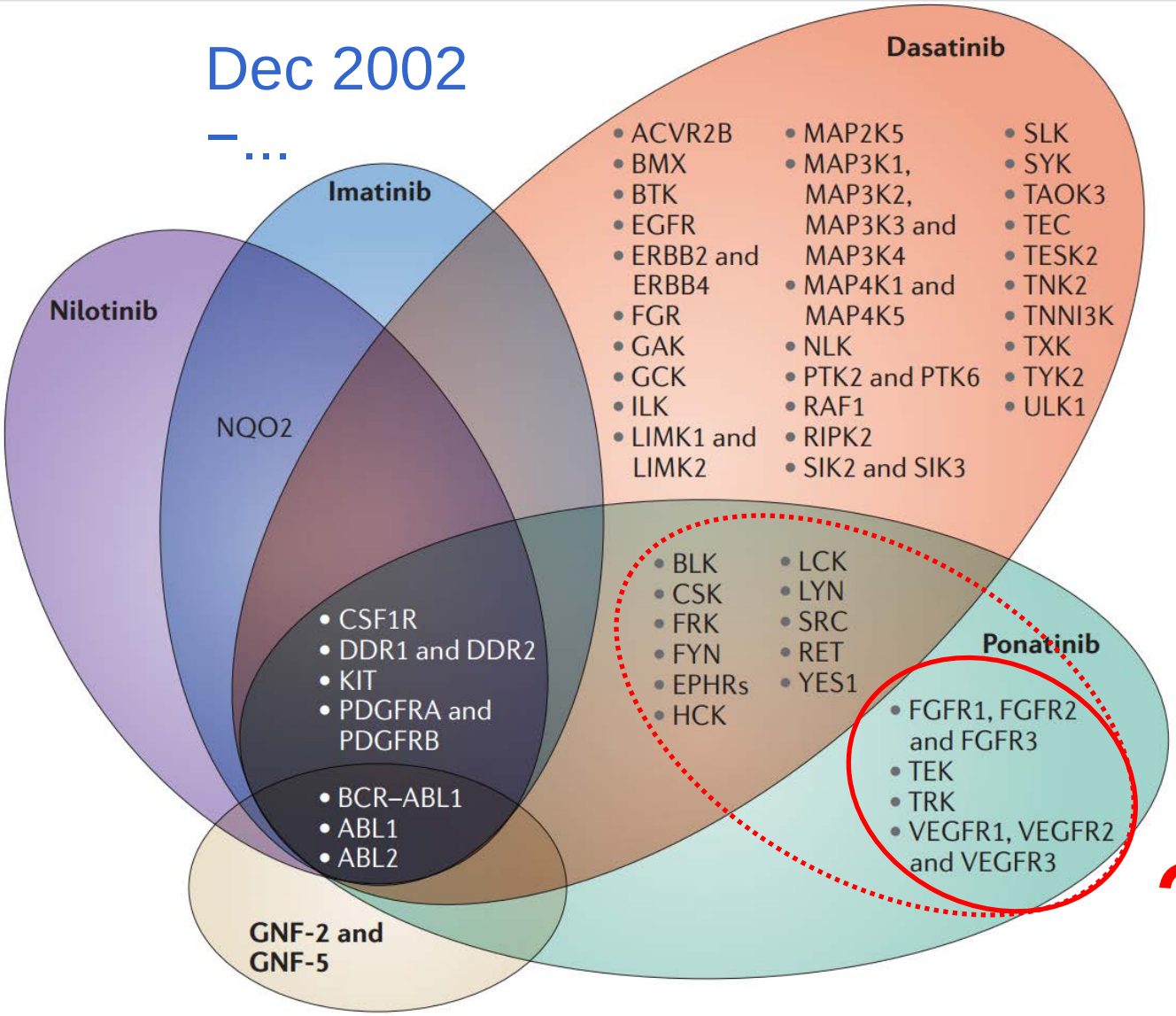
Complex Outcomes



Targets of BCR-ABL Inhibitors

Dec 2002 –…

Oct 2010 –…



Dec 2012 – Nov 2013

?

Abstract

When several (repeated, censored) events graded by severity and/or attribution need to be considered (e.g., hospitalization, myocardial infarction (MI), stroke, death), many strategies currently discussed either are not comprehensive enough or require unrealistic assumptions to be made, with often unpredictable consequences for validity.

We present a flexible approach to generalise the Gehan / Wei / Knuiman test to incorporate information about complex ordinal and hierarchical/factorial relationships between outcomes.

The proposed approach fulfills the ICH E9 criteria for “a predefined algorithm”, yet is flexible enough to objectively incorporates information about graded, repeated, and censored events for a wide range of complex diseases.

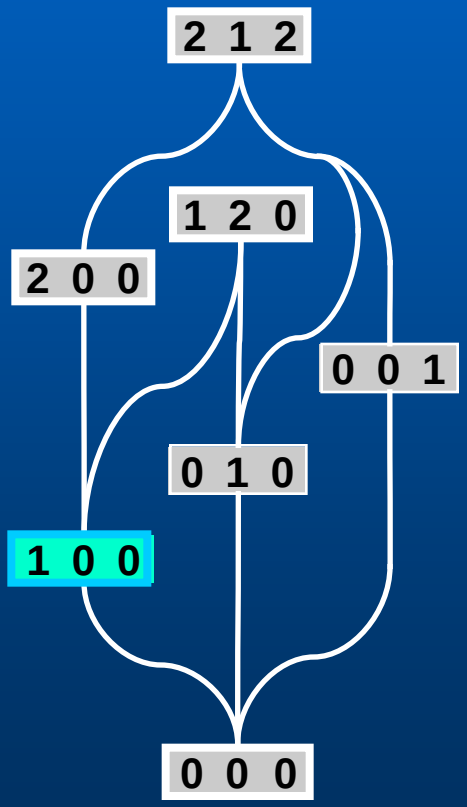
Overview

- U-statistics for multivariate data
  - Theory
  - Tools
  - Examples
- Increasing Information Content (IC)
  - AE / Response Structures
    - Censoring
    - Relatedness / Attribution
    - Grading by Severity
  - Genetic Structures
- “Gold mining” Phase III Trial Data

Multivariate Outcomes

Death
MI
Stroke

Partial Ordering Lattice



sCD25
IL1b
CD68/CD2

Efficacy
Cost
Safety

Environment
Society
Patient

SNP A / B / C

300 Years of Statistics:

From Notepads to Web Tools

John Arbuthnot (1667–1735)

1667: * Scotland
1692: translated HUYGENS' (1657) *De ratiociniis in ludo aleae* (1st English work on prob')
1696: MD from St Andrews Univ., taught mathematics
1704: Fellow of the *Royal Society*
1705: physician to Queen Anne
1714: founded the *Scriblerus Club* to satirise bad poetry and pedantry with T. PARNELL, J. GAY, A. POPE, and J. SWIFT (1729: *A modest proposal*).
1735: ✞ London

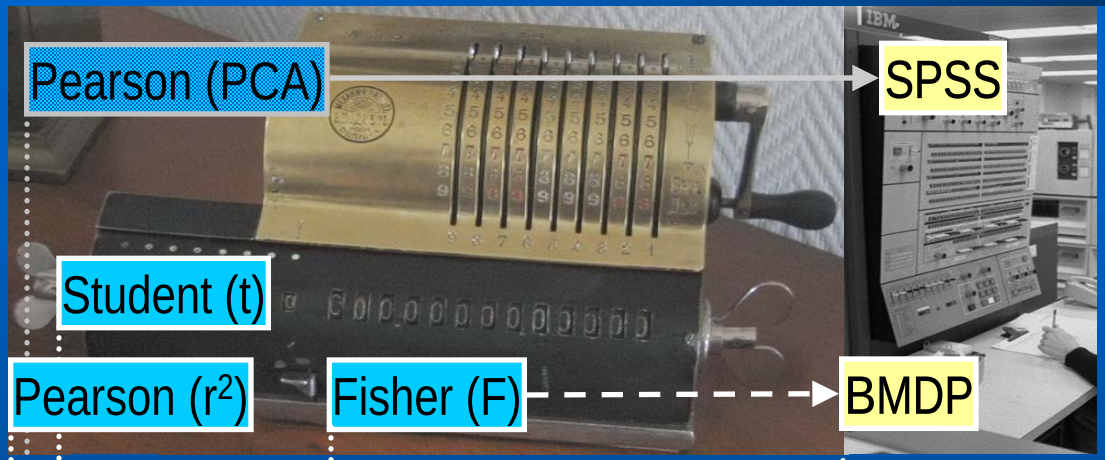
(1692) *On the laws of chance*

- discussing the slight excess of male births over female births
- demonstrating that **not chance** causes the sex ratio at birth
- the first application of probability to social statistics / medicine.

For „realistic" data (rounded, inexact, etc.), versions of this „**sign test**" were developed 250 years later ...

A Short History of Statistics

Linear Model



Pearson (PCA)

SPSS

Student (t)

Pearson (r^2)

Fisher (F)

BMDP

One of the attributes of ANOVA which ensured its **early popularity** was **computational elegance**. The structure of the additive model leds to **simple algebra** rather than **matrix calculations**.

en.wikipedial.org/wiki/Analysis_of_variance (2013-02-23)

W3 WXP/OSX

0 B 5 B 256 kB - 8 MB 16 MB 3 GB

1700 1900 1920 1940 1960 1980 2000 2020

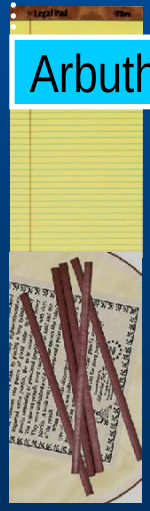
Arbuthnoth (sign test)

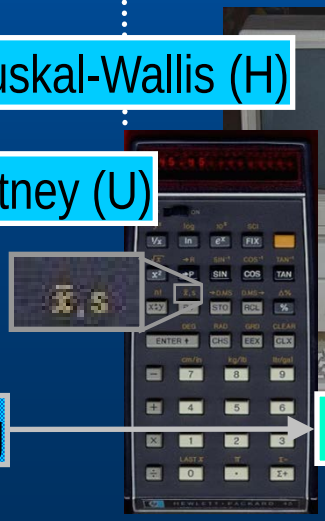
Kendall (τ)

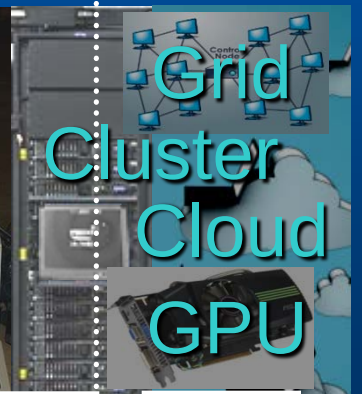
Kruskal-Wallis (H)

Mann-Whitney (U)

Hoeffding







Grid

Cluster

Cloud

GPU

...structures...

muStat

Computational Biostatistics

U-Statistics

One of the attributes of U-statistics, which ensures **biological validity** of results is its **conceptual simplicity**. The **matrix calculations** required, however, are **computationally demanding**.

U-Statistics for Multivariate Data:

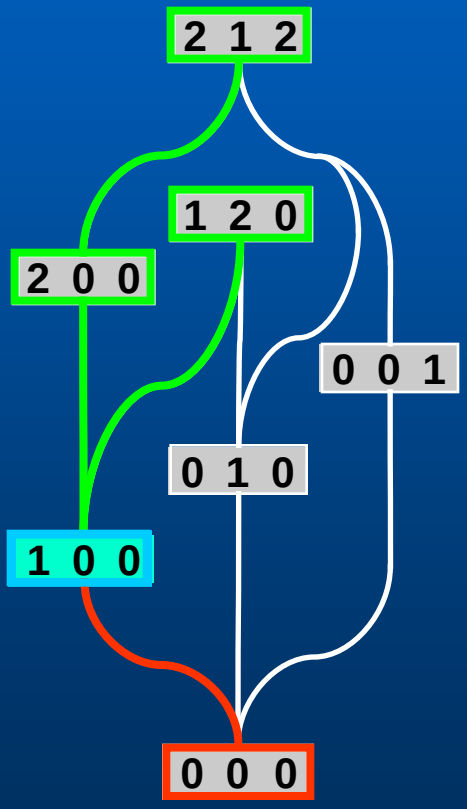
The Basics

A Class of Statistics

u scores are computed from *#(lower profiles)* *#(higher profiles)*.

u(1 0 0)
= 1−3 = −2

Partial Ordering Lattice



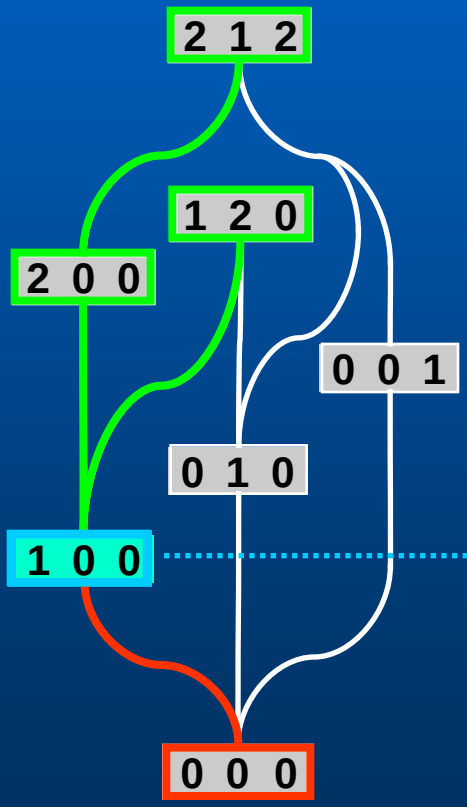
HOEFFDING W (1948) *AMS* 19:293

Algorithm (Deuchler, 1914):

u scores are computed as *#(lower profiles)* – *#(higher profiles)*.

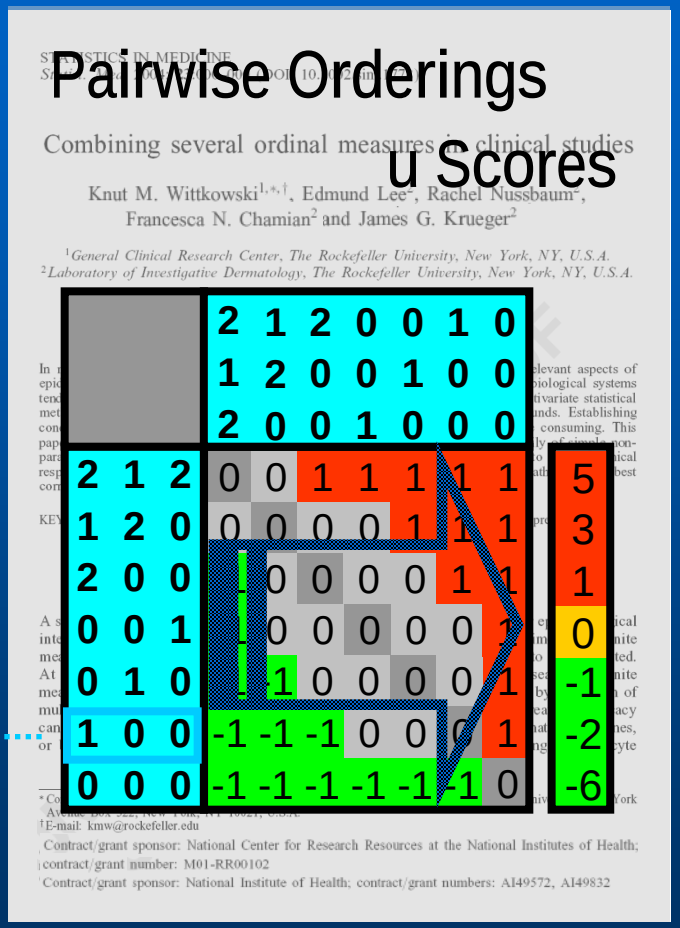
u(1 0 0) = 1−3 = −2

Partial Ordering Lattice



Pairwise Orderings

u Scores



WITTKOWSKI KM (2004) *Stat Med* 23

K-variate Censored Data

Let X_{skj} denote the time to event *k* for subject *j* in group *s*.

We observe $\left(\bar{X}_{skj}, \delta_{skj}\right)$, where $\bar{X}_{skj} = \begin{cases} \left(-\infty \;, \delta_{skj} = 1 \right) \\ \left(U_{skj} \left(< X_{skj} \right), \delta_{skj} = 1 \right) \\ \left(X_{skj} \;, \delta_{skj} = 0 \right) \end{cases} \Leftrightarrow \left(X^{-}_{skj}, X^{+}_{skj}\right) = \begin{cases} \left(-\infty \;, \infty \right) \\ \left(X_{skj}, \infty \right) \\ \left[X_{skj}, X_{skj}\right] \end{cases}.$

Wei, Knuiman (1987) define a score statistic

$$I_{WK}\left(\mathbf{X}_{1i} \le \mathbf{X}_{2j}\right) = \phi\left(\bar{X}_{1i}, d_{1i}; \bar{X}_{2j}, d_{2j}\right) = \begin{cases} 1 & \forall_k \; \bar{X}_{1ki} \le \bar{X}_{2kj}, \delta_{1ki} = 1 \\ 0 & \text{otherwise} \end{cases}$$

Censored Obs Only

"This type of order ... is rather strong. ... Several comparisons make no contribution." [p. 218]

Gehan (1965), Mantel (1966), Cox (1972), and Wei, Lachin (1984) (in the notation of Wittkowski (2004)) define a less "strong" score statistic

$$I\left(\mathbf{X}_{1i} \le \mathbf{X}_{2j}\right) = \phi\left(X^{-}_{1i}, X^{+}_{1i}; X^{-}_{2j}, X^{+}_{2j}\right) = \begin{cases} 1 & \forall_k \; X^{+}_{1ki} \le X^{-}_{2kj} \supset \delta_{1ki} = 0 \\ 0 & \text{otherwise} \end{cases}$$

No Overlap

The U statistic then is

$$u\left(\mathbf{X}_{1i}, \mathbf{X}_{2j}\right) = I\left(\mathbf{X}_{1i} \le \mathbf{X}_{2j}\right) - I\left(\mathbf{X}_{1i} \ge \mathbf{X}_{2j}\right) \in \left\{-1, 0, 1, NA\right\}$$

$$U_n = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} u\left(\mathbf{X}_{1i}, \mathbf{X}_{2j}\right) \;\square_{as.}\; N\left(0, \hat{\sigma}_0\right) \qquad s > 2, \text{ stratification: } W = \mathbf{u}'_{+} \mathbf{\Sigma}^{-}_{+} \mathbf{u}_{+} \;\square_{as.}\; \chi^2_{s-1}(0)$$

Wittkowski 1988

"The likelihood of indefiniteness ... will increase as the dimensionality of the data increases."

RUH.EDBS U Score Spreadsheet

ID	TTI	Chol	Lath	Camp	Sito	rChol	rLath	Camp	rSito	TG	0	0	0	0	0	0	-1	1	1	0	#<	#>	U
623	212	0.37	0.66	0.15	0.16	1.77	3.10	0.73	0.78	79	?	?	?	?	?	?	2	10	7	?	0	27	-27
1944	159	0.26	0.38	0.11	0.11	1.64	2.41	0.66	0.71	108	?	?	?	?	?	?	12	4	2	?	0	25	-25
585	191	0.32	0.57	0.14	0.17	1.67	2.98	0.73	0.90	75	?	?	?	?	?	?	4	10	14	?	1	23	-22
1446	129	0.23	0.34	0.09	0.13	1.81	2.62	0.71	1.01	93	?	?	?	?	?	?	10	6	20	?	0	19	-19
1565	163	0.26	0.36	0.15	0.13	1.61	2.20	0.93	0.78	137	?	?	?	?	?	?	16	24	7	?	2	18	-16
549	216	0.28	0.58	0.19	0.24	1.31	2.70	0.87	1.10	161	?	?	?	?	?	?	8	22	24	?	2	17	-15
1513	148	0.29	0.25	0.13	0.11	1.98	1.70	0.86	0.76	57	?	?	?	?	?	?	35	19	4	?	1	13	-12
66	249	0.52	0.43	0.21	0.21	2.07	1.73	0.84	0.85	133	?	?	?	?	?	?	30	16	12	?	2	14	-12
90	228	0.39	0.36	0.17	0.19	1.69	1.59	0.76	0.81	105	?	?	?	?	?	?	38	14	10	?	2	12	-10
560	169	0.26	0.31	0.12	0.24	1.54	1.85	0.73	1.42	261	?	?	?	?	?	?	24	10	38	?	4	9	-5
446	162	0.3	0.33	0.16	0.20	1.83	2.04	0.99	1.25	174	?	?	?	?	?	?	22	32	30	?	6	11	-5
519	136	0.27	0.33	0.15	0.21	1.96	2.39	1.07	1.58	198	?	?	?	?	?	?	14	40	44	?	5	7	-2
424	239	0.44	0.52	0.25	0.30	1.82	2.18	1.06	1.27	119	?	?	?	?	?	?	20	38	32	?	6	8	-2
1082	148	0.27	0.22	0.08	0.32	1.81	1.49	0.53	2.18	139	?	?	?	?	?	?	40	2	62	?	0	1	-1
1573	213	0.48	0.35	0.23	0.26	1.75	1.27	0.86	0.96	111	?	?	?	?	?	?	50	19	16	?	6	7	-1
523	163	0.23	0.29	0.16	0.22	1.43	1.81	1.00	1.33	90	?	?	?	?	?	?	26	36	36	?	7	7	0
17	125	0.20	0.22	0.12	0.20	1.60	1.76	0.99	1.56	249	?	?	?	?	?	?	28	32	42	?	8	7	1
1140	214	0.45	0.35	0.34	0.23	2.22	1.71	1.68	1.11	76	?	?	?	?	?	?	32	50	26	?	7	6	1
577	169	0.29	0.37	0.21	0.29	1.70	2.19	1.22	1.70	81	?	?	?	?	?	?	18	45	48	?	7	5	2
1502	181	0.47	0.50	0.40	0.36	2.57	2.76	2.19	2.00	93	?	?	?	?	?	?	6	64	58	?	2	0	2
550	120	0.23	0.16	0.11	0.14	1.92	1.29	0.94	1.18	136	?	?	?	?	?	?	48	27	28	?	9	5	4
384	179	0.37	0.13	0.19	0.17	2.09	0.75	1.08	0.97	49	?	?	?	?	?	?	64	42	18	?	8	0	8
639	257	0.48	0.25	0.24	0.27	1.86	0.96	0.94	1.06	152	?	?	?	?	?	?	60	27	22	?	9	1	8
1874	157	0.39	0.21	0.28	0.21	2.45	1.34	1.78	1.32	41	?	?	?	?	?	?	43	54	34	?	12	3	9
1414	190	0.45	0.25	0.30	0.28	2.39	1.34	1.61	1.49	90	?	?	?	?	?	?	43	48	40	?	13	4	9
561	109	0.23	0.11	0.11	0.19	2.09	1.00	0.99	1.71	181	?	?	?	?	?	?	58	32	50	?	14	1	13
1127	131	0.32	0.22	0.25	0.26	2.44	1.70	1.90	1.99	42	?	?	?	?	?	?	35	58	56	?	16	1	15
554	153	0.35	0.17	0.19	0.24	2.28	1.09	1.22	1.59	109	?	?	?	?	?	?	56	45	46	?	17	1	16
1220	140	0.33	0.19	0.29	0.29	2.36	1.33	2.05	2.09	50	?	?	?	?	?	?	46	62	60	?	20	0	20
1862	172	0.44	0.21	0.31	0.30	2.55	1.24	1.79	1.76	75	?	?	?	?	?	?	53	56	52	?	21	1	20
605	200	0.47	0.25	0.41	0.39	2.33	1.24	2.03	1.95	59	?	?	?	?	?	?	53	60	54	?	22	0	22
568	169	0.42	0.13	0.29	0.52	2.45	0.78	1.69	3.10	93	?	?	?	?	?	?	62	52	64	?	24	0	24

Sort By ID

Sort By UScr

Disclaimer: Until knowledge of the study design (protocol based meta data) will become available in the hospital data base, number of groups, group names (if any), subject names, group-subject association, variable structure (e.g., pairs) and format, data, transformations, and test statistics cannot be derived or suggested, so that the user is solely responsible for reentering or choosing correct and appropriate data, meta data, and functions.

kmw@rockefeller.edu
Wittkowski KM et al.
Statistics in Medicine (in press)

rLath
rCamp
rSito
?

pat. pending

Contact
Help
µStat
Start New Job
List Jobs
Utilities
WISDOM
BERD DashBoard
Study Design/Form Design
REDCap
Export to REDCap and Epi Info
Import from REDCap
Open REDCap
Internal
Script Manager
List All Jobs
Server Stats
Clean Up
Testing
Don't touch

Examples

Example 1: Scoring Multivariate Data

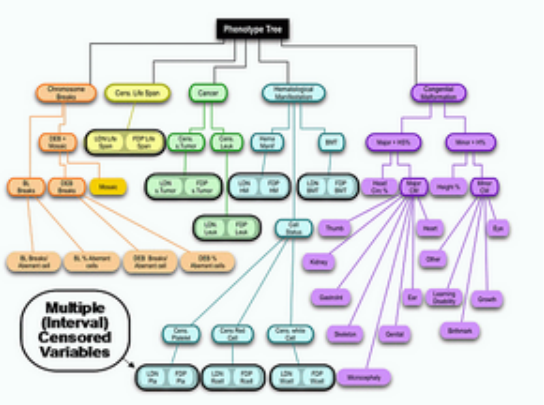
One of the simplest applications of µStat is to score a set of p rofiles (multivariate data). See Table I in Wittkowski et al. (2004) *Statistics in Medicine* for details. For this small data set, the results could be obtained by downloading the spreadsheet U32x10N0. Variables are selected by placing a "polarity" (-1: less is better, 0: ig nore, 1: more is better) on top of the turquoise columns. When uploading the data to muStat.rockefeller.edu, the Parameter fi le is currently ignored when merely scoring data (Parameter file/Data file/Results).

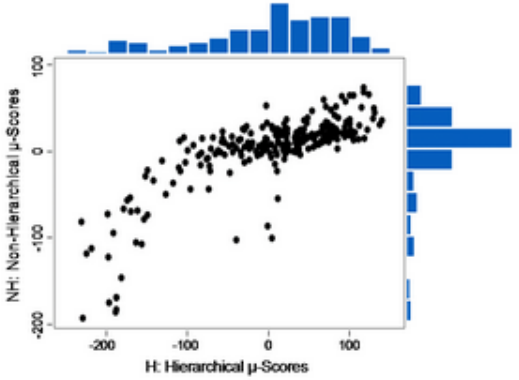
Example 2: Selecting Relevant Subsets of Variables

The output from Example 1 could then be used to find a combination of variables whose scores correlate bes t with the outcome. The data below are, again, from Table I in Wittkowski et al. (2004), the the basis of Table III. For a given combination of variables, the results can be obtained after downloading the spre adsheet U32x10N1, and placing the u-scores obtained in Example 1 into the column &quo t;Ref" to the left of the data (small differences ma y occur due to rounding errors). Even for only six variables, the nu mber of possible polarized combinations is daunting. The co plete list of results (including those given in Table III, can be obtained by uploading parameters and data to muStat.rockefeller. (Parameter file / Data file / Results).

Example 3: Including multiple (singly/doubly interval) censored variables into a factor structure

As the number of variables increases, the proportion of pairwise orderings that can not be decided (shown as grey cells in the "chessboard" of the above spreadsheets) tends to increase. This loss of information content ca n be reduced, if it is known, which variables belong to the same (sub-) factor. With µ-Scores, several censored variabl es can be incorporated into a factor structure. In Morales J et al. (2008) *Stat Appl Genet Mol Biol*, hierarchical scoring of Fanconi anemia severity indicators resulted in a more informative FA phenotype score (FAPS). The figu re below is based on this Data file, where the header rows selected are either "Pht.1/Pht .2/Pht.1" (hierarchical) or "Pht" (nonhierarchical); the limits of interval-censored variables are indicated by the row heade r "lvl".

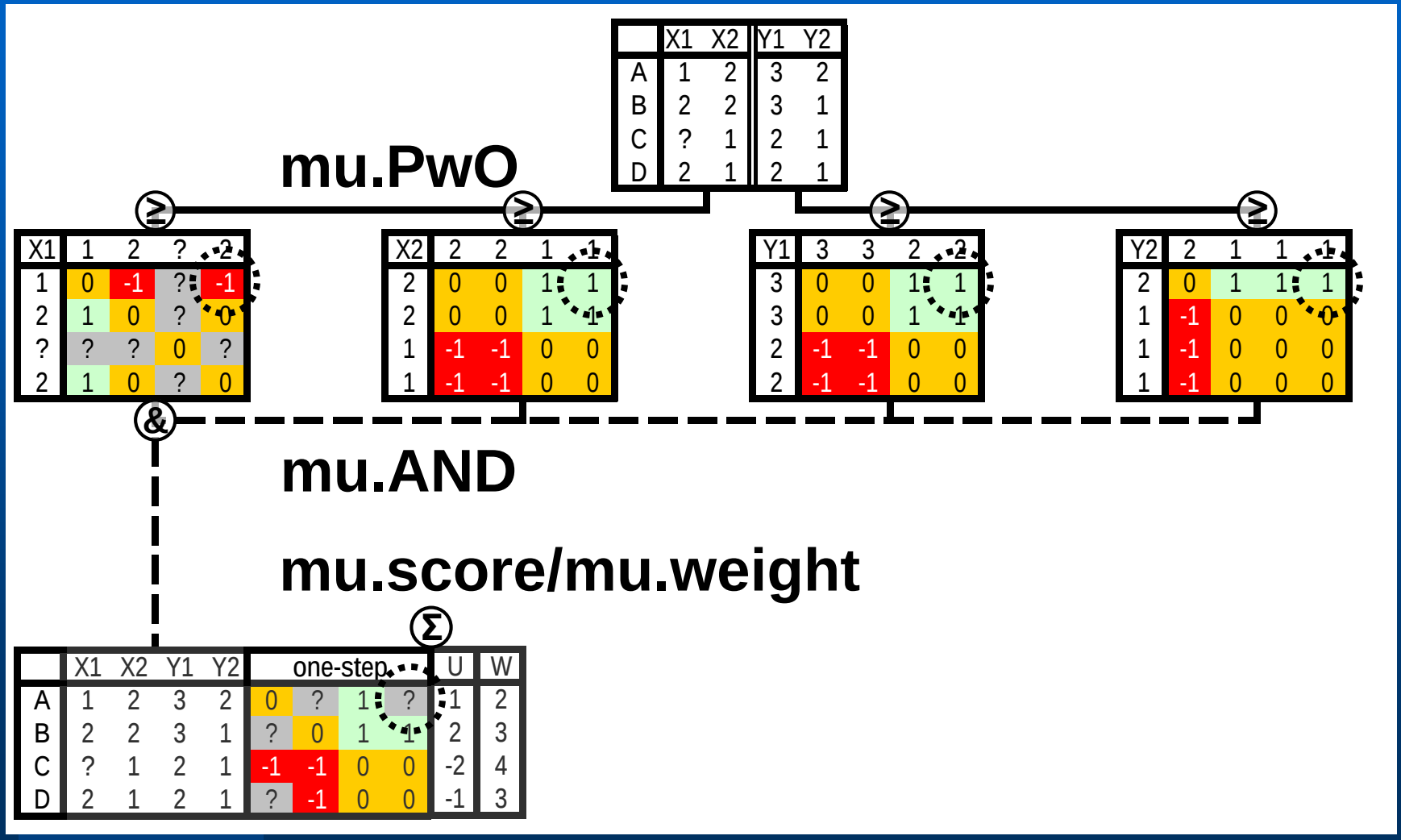




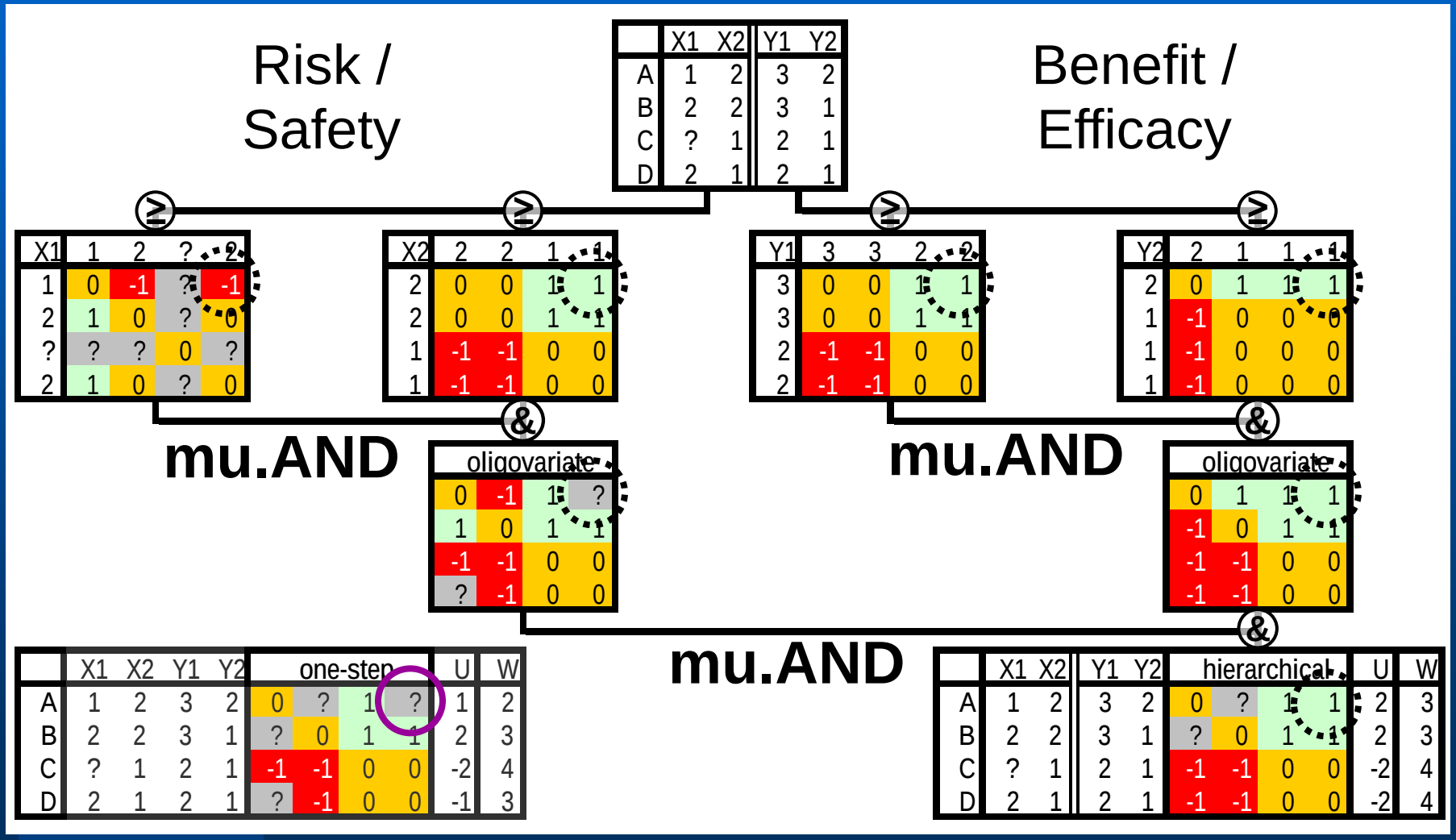
U-Statistics for Multivariate Data:

Structured Variables

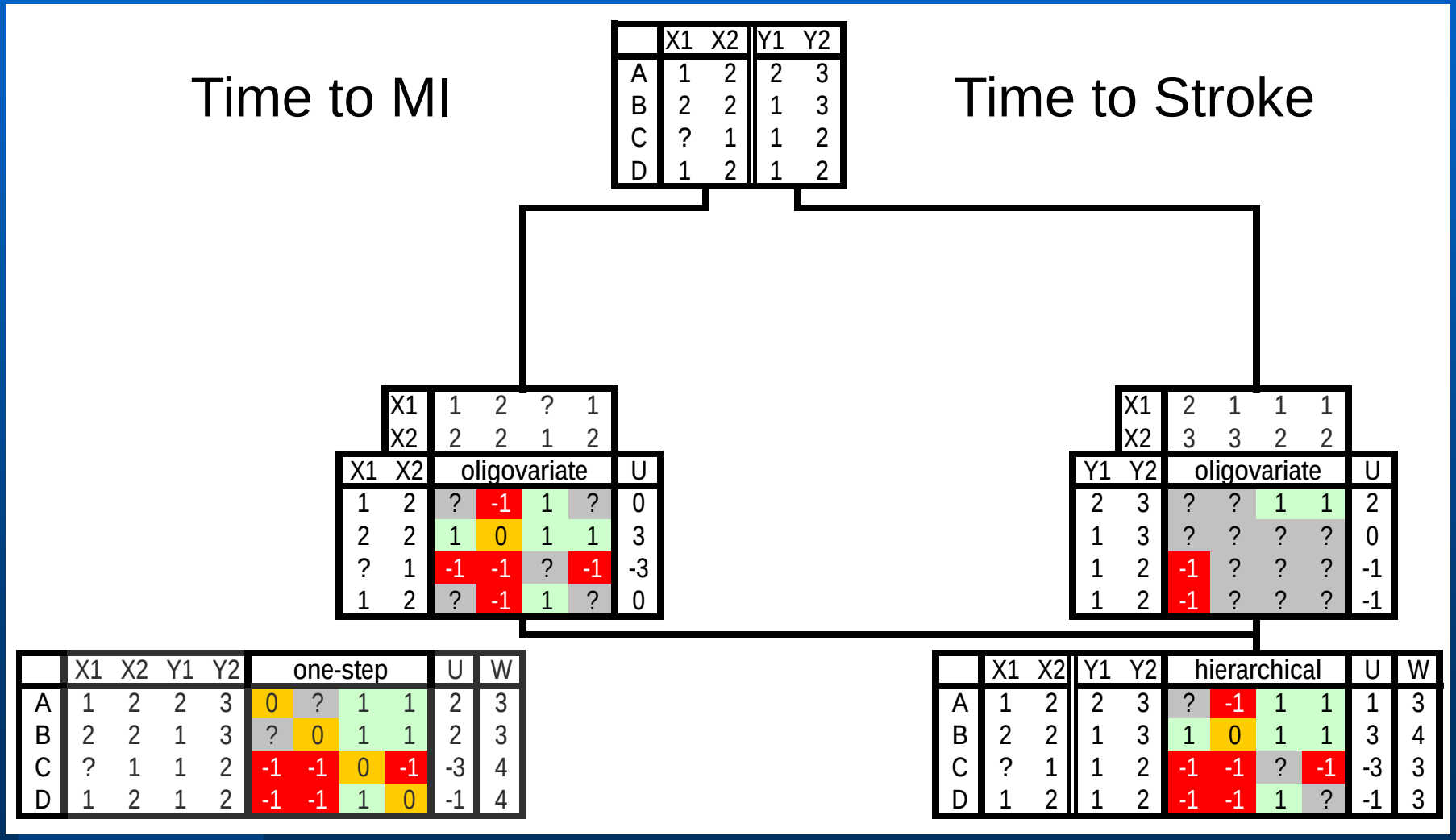
#Variables vs Information content



Hierarchy Increases Information



Multiple Interval Censored Data



WEI KNUIMAN (1987) *Austr J Stat* 29: 214

U-Statistics for Multivariate Data:

Tools

S/R Functions

- **mu.GE**

  pairwise comparison

- **mu.AND**

  combine pairwise comparisons allowing for hierarchical formula

- **mu.Sums**

  generate scores and weights from (combination of) pairwise comparisons

```
mu.GE <- function(x, y=x) { …
  if (length(y)>1) apply(rbind(x,y),2,mu.GE,nrow(x))
  else as.numeric(NAtoZer(outer(x[1:y],x[-(1:y)],">="))) }

mu.AND <- function(GE, frml=NULL) { …
  if (is.null(frml)) {
    GE  <- sq.array(GE), AND <- GE[,,1]^0, nNA <- AND[,1]*0
    for (i in 1:dim(GE)[3]) {
      nNA <- nNA + diag(GEi <- GE[,,i])
      AND <- AND * (GEi + (1-GEi)*(1-t(GEi))) }
    return(as.numeric(AND * ((c(nNA)%o%c(nNA))>0))) }
  else # … deal with the formula …
  return(tmp[-1,1]) }

mu.Sums <- function(GE, dsgn=1) {
  ICW <- function(GE,dsgn) {
    wgt <- colSums(GE|t(GE))
    sqrt(wgt*(wgt>1)/if (dsgn!=1) colSums(dsgn) else nrow(GE)}
  GE   <- sq.matrix(GE)
  wght <- ICW(GE,dsgn)
  list (
    score = (rowSums(GE) - colSums(GE)) *ifelse(wght==0,NA,1),
    weight  = wght) }
```

mu.test

The new function '**mu.test**' integrates / extends
(at twice the speed)
several well-known tests:

	Conditions	Granularity	Replications	Blocks
mcnemar.test	2	2	>1	1
SMN.test	2	2	>1	3
CA.test (new)	2	3	>1	1
wilcox.test	2	>1	>1	1
kruskal.test	>1	>1	>1	1
friedman.test	>1	>1	1	>1
mu.test	**>1**	**>1**	**≥0**	**>1**

CRAN

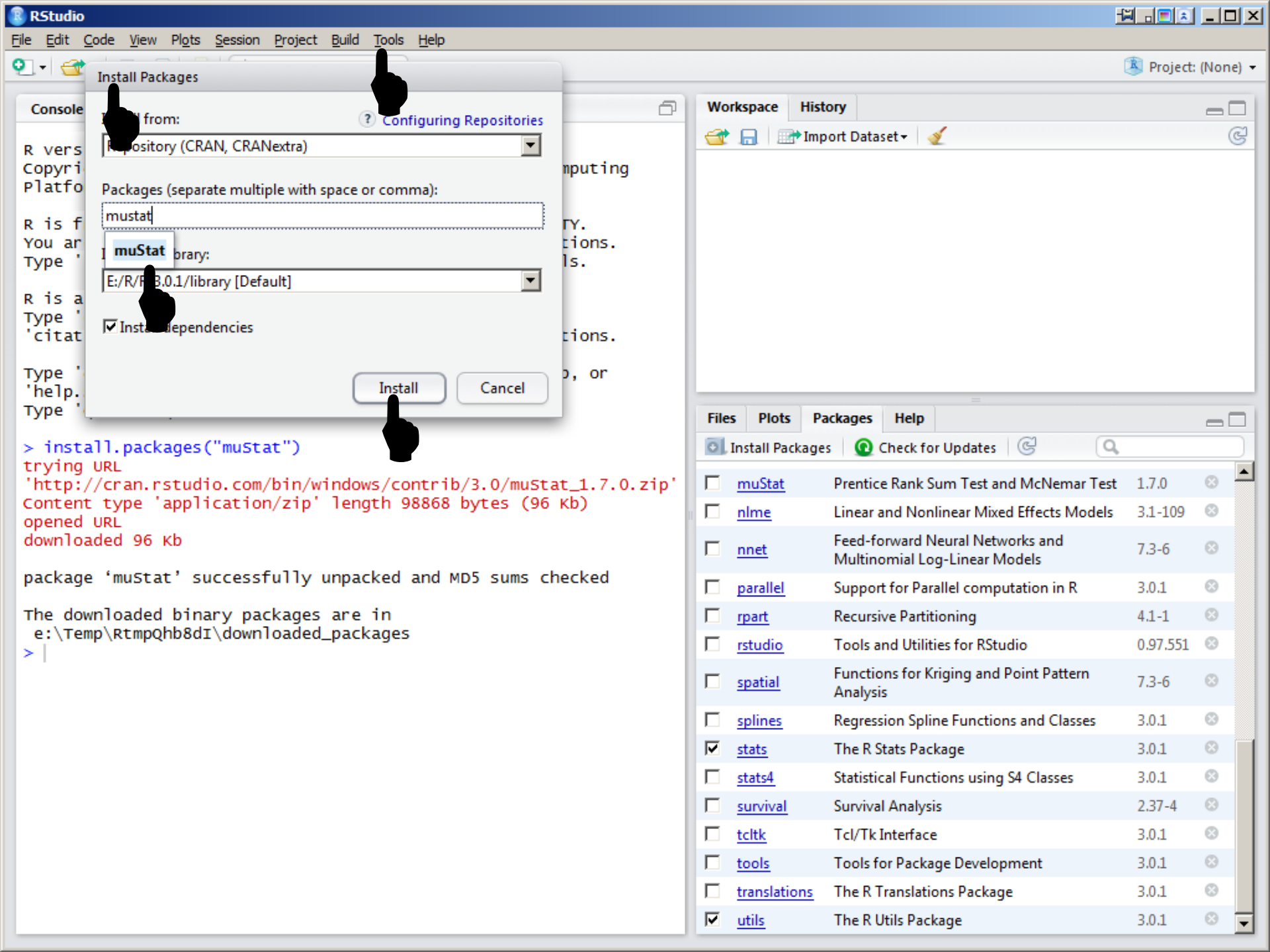
http://cran.r-project.org/index.html

CSAN

http://csan.insightful.com/Default.aspx

mu.test()


```
mu.test <- function(
    y,                          # data   (NA ok)
    groups,                     # groups (unbalanced ok)
    blocks  = NULL,             # blocks (unequal size ok)
    score   = "rank",           # NULL: already scored
    blkwght = "prentice“,       # block weights
                                # Wittkowski (1988) JASA

    alternative = "two.sided",      # wilcox only
    mu          = 0,                # wilcox only
    paired      = FALSE,            # wilcox only
    exact       = NULL,             # wilcox only
    correct     = NULL              # wilcox only
)
```


RStudio
File Edit Code View Plots Session Project Build Tools Help
Project: (None)
Install Packages
from:
Configuring Repositories
Repository (CRAN, CRANextra)
Packages (separate multiple with space or comma):
mustat
muStat
E:/R/R-3.0.1/library [Default]
Install dependencies
Install
Cancel
Workspace History
Import Dataset
Files Plots Packages Help
Install Packages
Check for Updates
> install.packages("muStat")
trying URL
'http://cran.rstudio.com/bin/windows/contrib/3.0/muStat_1.7.0.zip'
Content type 'application/zip' length 98868 bytes (96 Kb)
opened URL
downloaded 96 Kb
package 'muStat' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
e:\Temp\RtmpQhb8dI\downloaded_packages
muStat Prentice Rank Sum Test and McNemar Test 1.7.0
nlme Linear and Nonlinear Mixed Effects Models 3.1-109
nnet Feed-forward Neural Networks and Multinomial Log-Linear Models 7.3-6
parallel Support for Parallel computation in R 3.0.1
rpart Recursive Partitioning 4.1-1
rstudio Tools and Utilities for RStudio 0.97.551
spatial Functions for Kriging and Point Pattern Analysis 7.3-6
splines Regression Spline Functions and Classes 3.0.1
stats The R Stats Package 3.0.1
stats4 Statistical Functions using S4 Classes 3.0.1
survival Survival Analysis 2.37-4
tcltk Tcl/Tk Interface 3.0.1
tools Tools for Package Development 3.0.1
translations The R Translations Package 3.0.1
utils The R Utils Package 3.0.1

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function

Project: (None)

Console ~/

```
R version 3.0.1 (2
Copyright (C) 2013
Platform: x86_64-w

R is free software
You are welcome to
Type 'license()' o

R is a collaborati
Type 'contributors
'citation()' on ho

Type 'demo()' for
'help.start()' for
Type 'q()' to quit

> install.packages
trying URL
'http://cran.rstud
Content type 'appl
opened URL
downloaded 96 Kb

package 'muStat' s

The downloaded bir
 e:\Temp\RtmpQhb8d
>
```

Workspace History

Import Dataset

Files Plots Packages Help

R: Prentice (Friedman/Wilcoxon/Kruskal) Rank Sum Test

Find in Topic

Prentice (Friedman/Wilcoxon/Kruskal) Rank Sum Test

Description

Performs a generalized Friedman rank sum test with replicated blocked data or, as special cases, a Kruskal-Wallis rank sum test on data following a one-way layout or a Wilcoxon rank sum test following a one-way layout with only two groups.

Usage

```
prentice.test(y, groups, blocks = NULL,
  score = "rank", blkwght = "prentice", condvar = TRUE,
  alternative = "two.sided", mu = 0, paired = FALSE,
  exact = NULL, correct = FALSE, df = -1, warn = 0, optim = TRUE)
mu.wilcox.test(y, groups, blocks = NULL, score = "rank",
  paired = FALSE, exact = TRUE, correct = TRUE, ...)
mu.kruskal.test(y, groups, blocks, ... )
mu.friedman.test(y, groups, blocks, ... )
```

Arguments

`y`	a numeric vector of data values, `NA`s are used to compute block weights with incomplete block data (see `blkwght`), but will be removed with one-way designs. Blocks with observations in only one group will be removed. `Inf`s are allowed, and are not removed as they are rankable. Required.
`groups`	factor or category object of the same length as `y`, giving the group (treatment) for each corresponding element of `y`. `NA`s are allowed and observations with `NA` in `groups` will be used for scoring other observations, treating them as if they were observations in an additional (fictional) group. If not a factor or category object, it will be coerced to one. Required.
`blocks`	factor or category object of the same length as `y`, giving the block membership for each corresponding element of `y`. Observations with `NA` in `blocks` will be removed. If not a factor or category object, it will be coerced to one.
`score`	character or function object, giving the score function to be used. If `NULL`, `y` is assumed to already have been scored, e.g., by marginal likelihood scores (Wittkowski 1992) or u-scores (Wittkowski 2004) for multivariate data. If a function, it is

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function

Project: (None)

Console ~/

```
R version 3.0.1 (2
Copyright (C) 2013
Platform: x86_64-w

R is free software
You are welcome to
Type 'license()' o

R is a collaborati
Type 'contributors
'citation()' on ho

Type 'demo()' for
'help.start()' for
Type 'q()' to quit

> install.packages
trying URL
'http://cran.rstud
Content type 'appl
opened URL
downloaded 96 Kb

package 'muStat' s

The downloaded bir
 e:\Temp\RtmpQhb8c
>
```

Workspace History

Import Dataset

Files Plots Packages Help

R: Prentice (Friedman/Wilcoxon/Kruskal) Rank Sum Test

Find in Topic

Prentice (Friedman/Wilcoxon/Kruskal) Rank Sum Test

Description

Performs a generalized Friedman rank sum test with replicated blocked data or, as special cases, a Kruskal-Wallis rank sum test on data following a on

Usage

```
prentice.t
  score =
  alternat
  exact =
mu.wilcox.
  paired =
mu.kruskal
mu.friedma
```

Arguments

y	
groups	
blocks	
score	

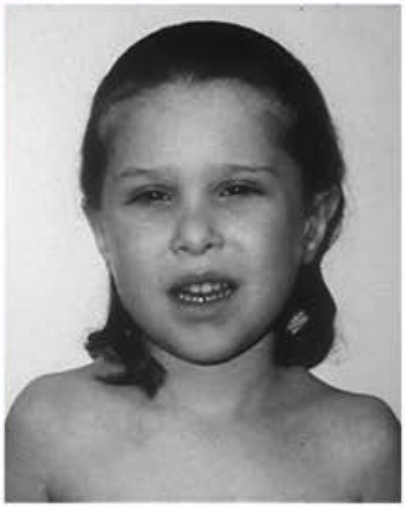
EXAMPLES:

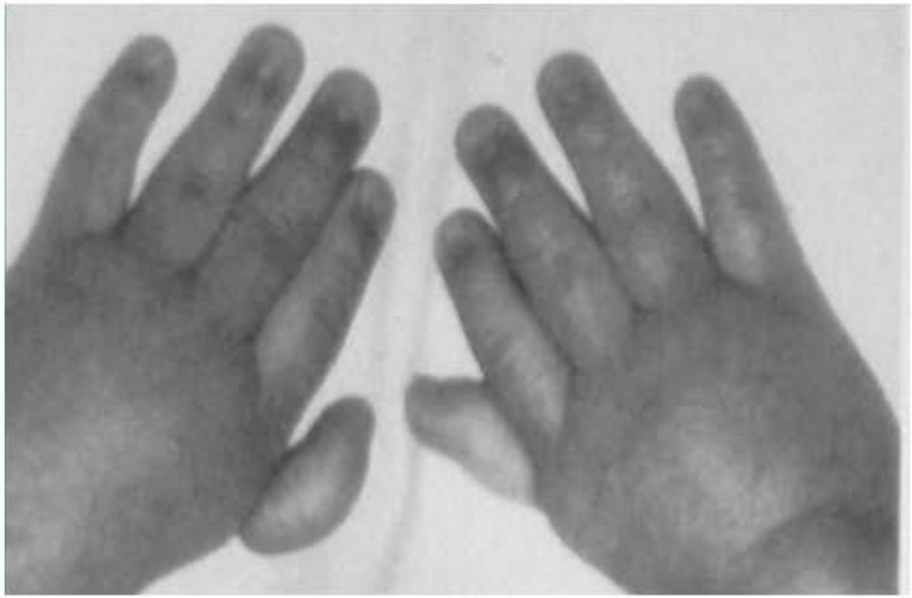
```
# Special case: Kruskal Wallis

# Data from Hollander and Wolfe (1973), p. 116
holl.y <- c(2.9,3.0,2.5,2.6,3.2,3.8,2.7,4.0,2.4,2.8,3.4,3.7,2.2,2.0)
holl.grps <- factor(c(1,1,1,1,1,2,2,2,2,3,3,3,3,3),
     labels=c("Normal Subjects","Obstr. Airway Disease","Asbestosis"))
kruskal.test(holl.y, holl.grps)
prentice.test(holl.y, holl.grps)

# Now suppose the data is in the form of a table already,
# with groups in columns; note this implies that group
# sizes are the same.
tab.data <- matrix(c(.38,.58,.15,.72,.09,.66,.52,.02,.59,.94,
     .24,.94,.08,.97,.47,.92,.59,.77), ncol=3)
tab.data

# Generate 'y' and 'groups':
y2 <- as.vector(tab.data)
gr <- factor(as.vector(col(tab.data)))    # Groups are columns
kruskal.test(y2, gr)
prentice.test(y2, gr)
```

So Much for Theory

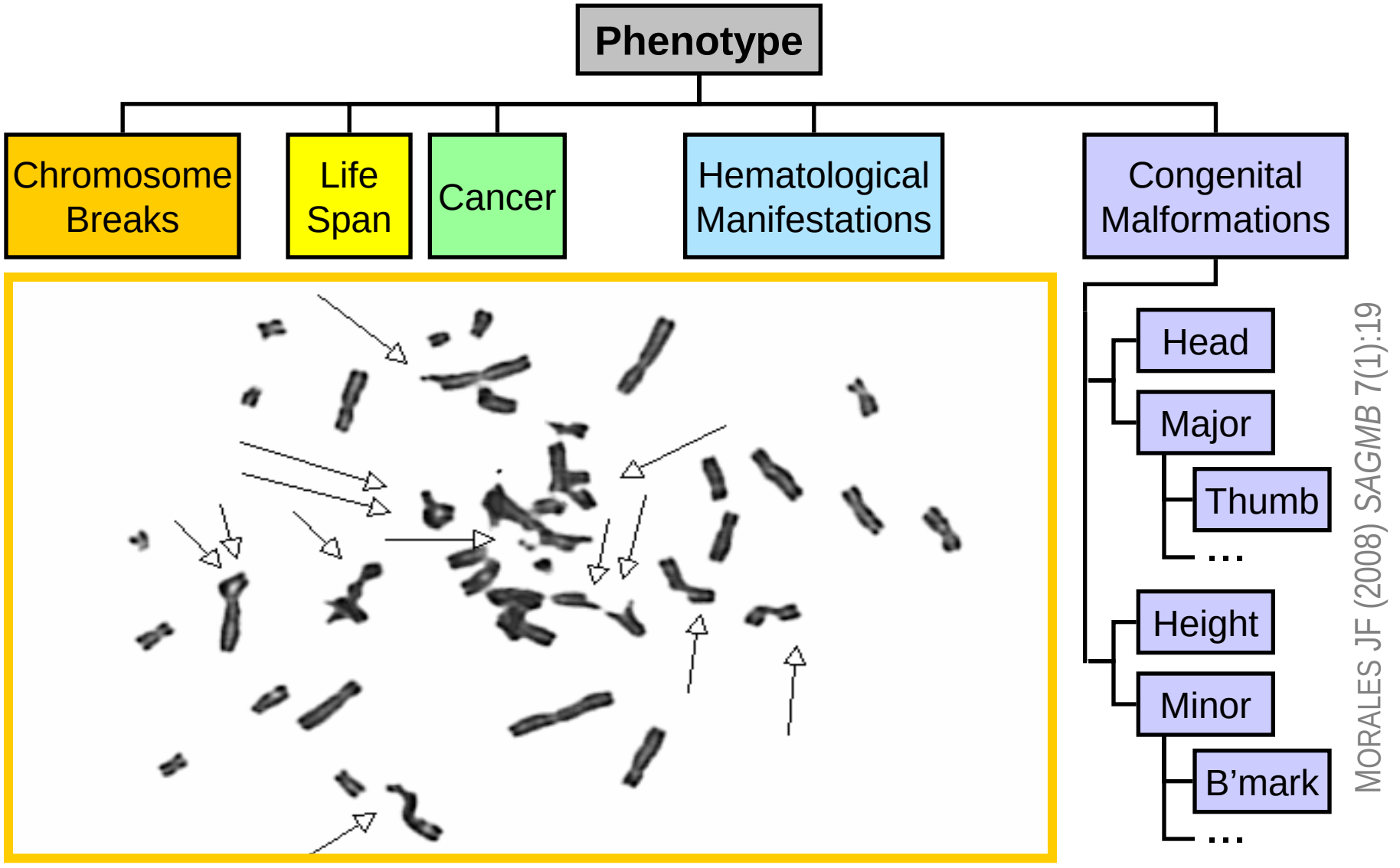
Complex Phenotypes w/censoring

Fanconi Anemia Patients

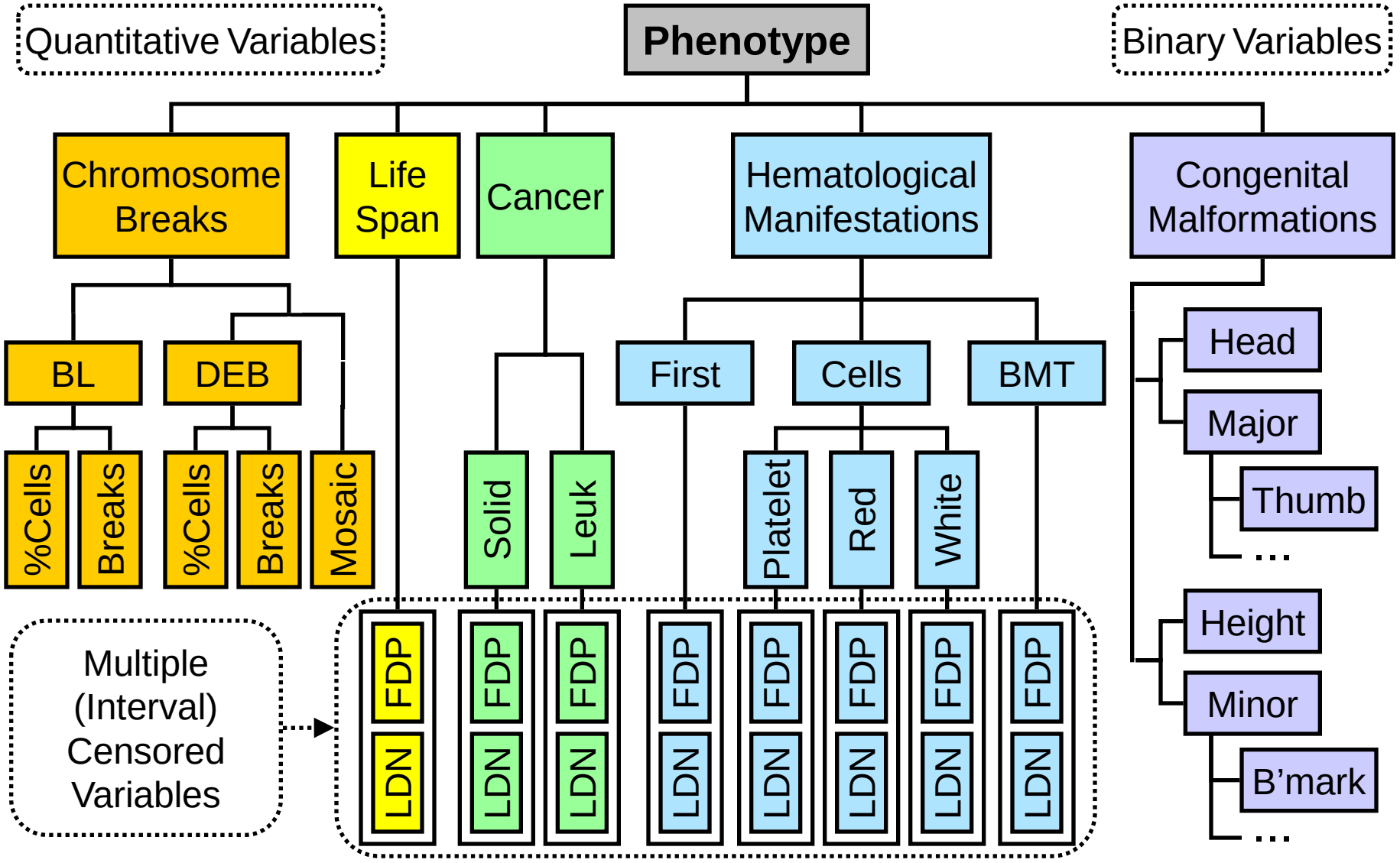





µPhene: Complex Phenotype

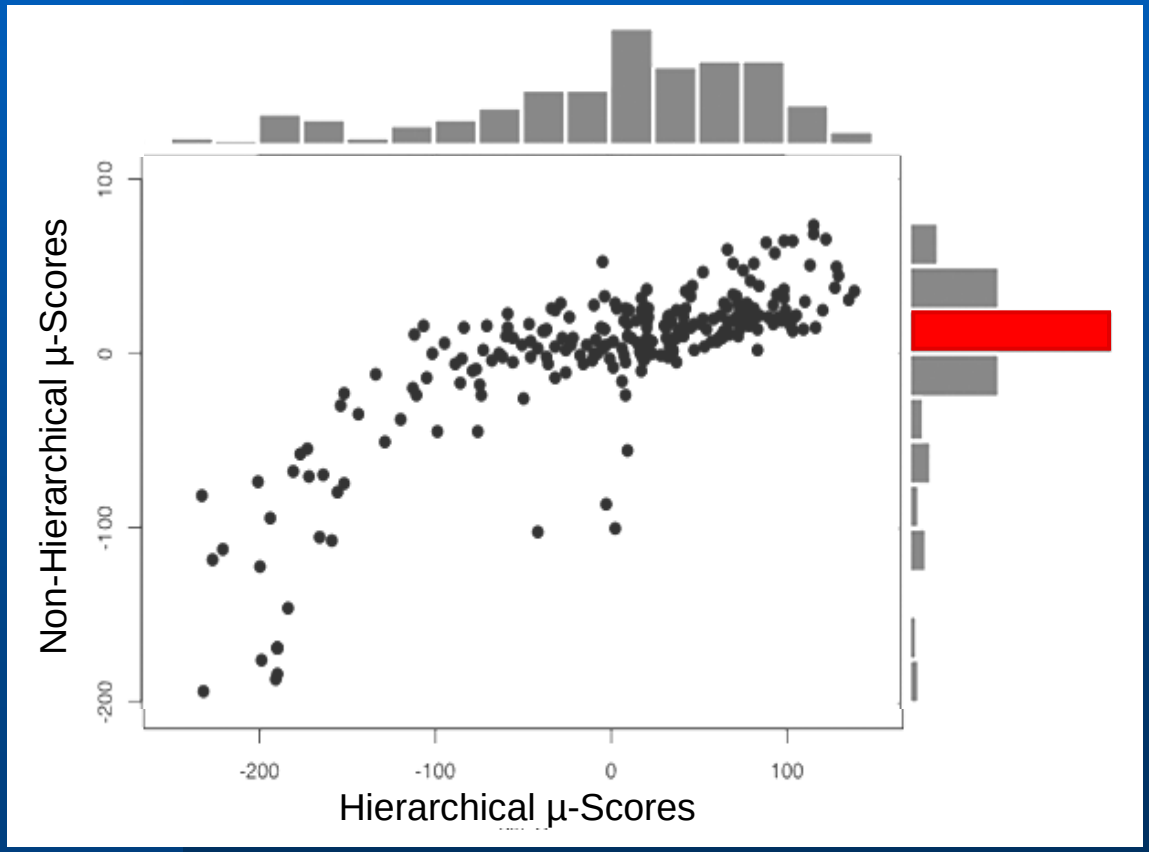


Fanconi Anemia: Complex Phenotype

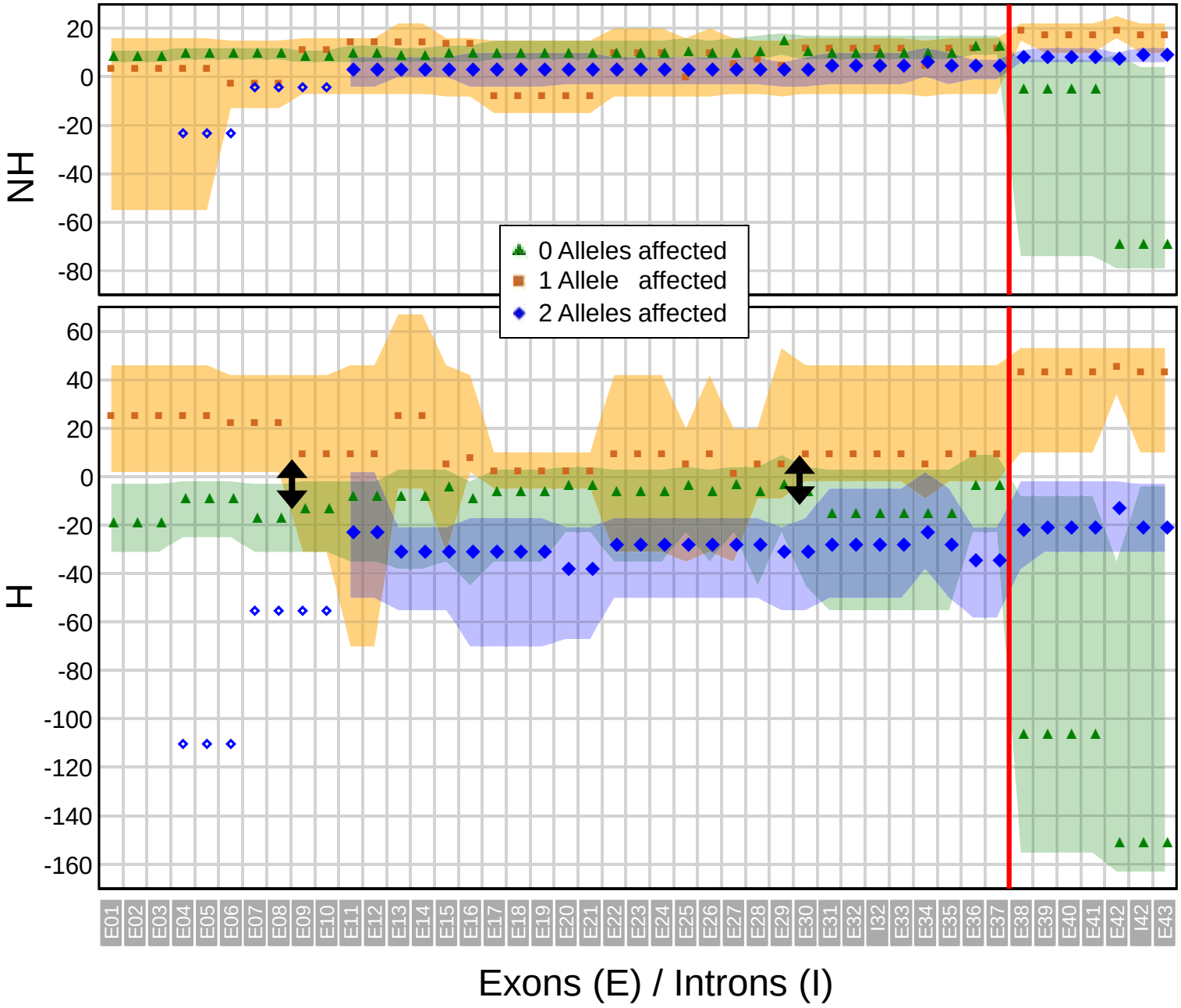


Advantages of Hierarchical μ-Scores

More Information Content
More Balanced Sensitivity



Larger Signal/Noise Ratio
Better Discrimination

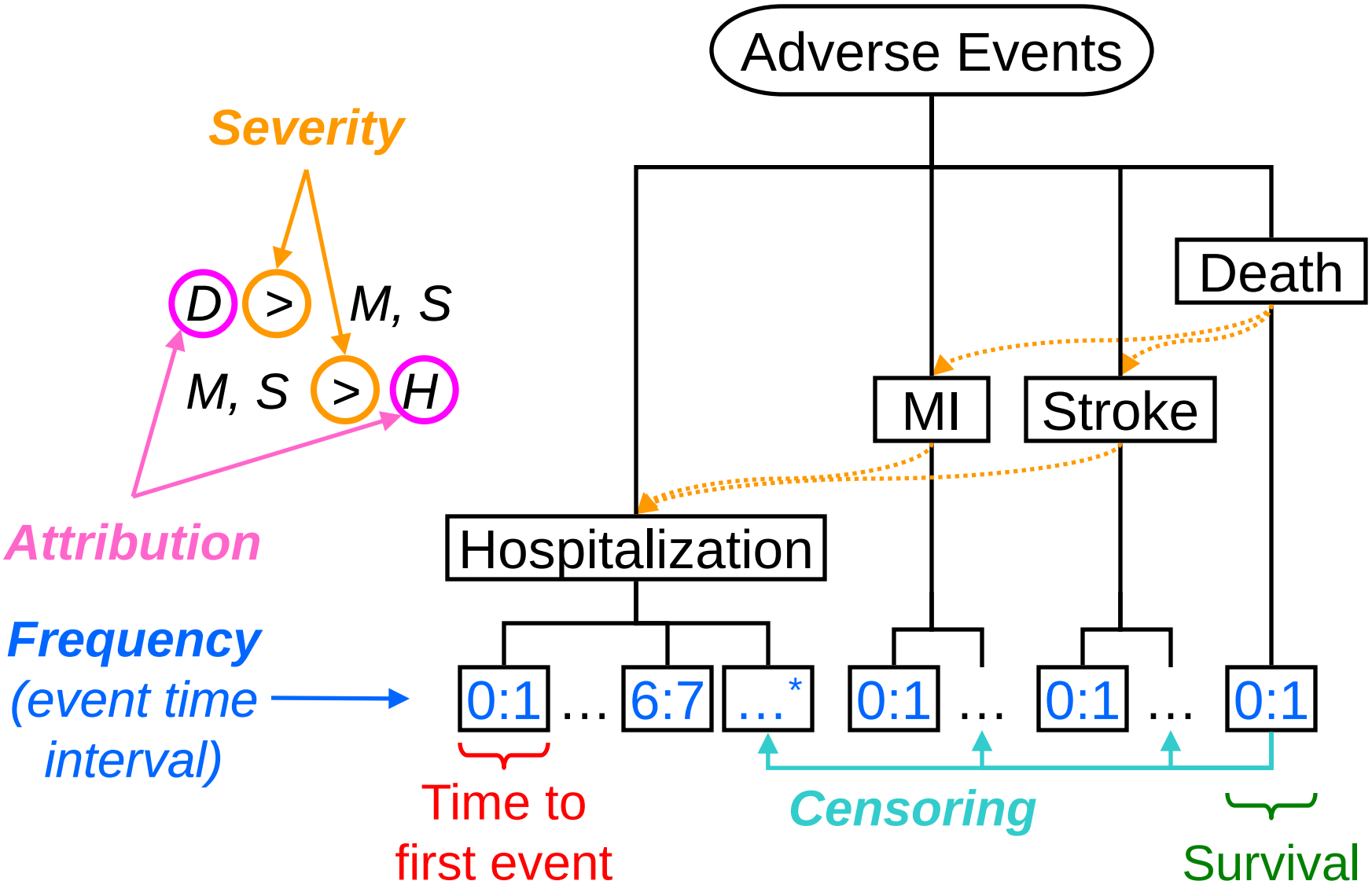


MORALES JF (2008) SAGMB 7(1):19

U-Statistics for Multivariate Data:

Graded Variables

µPhene Combines Complex Outcomes



Graded Variables

Let

- benign,
- severe, and
- grave

Graded Variables

Severity of adverse events

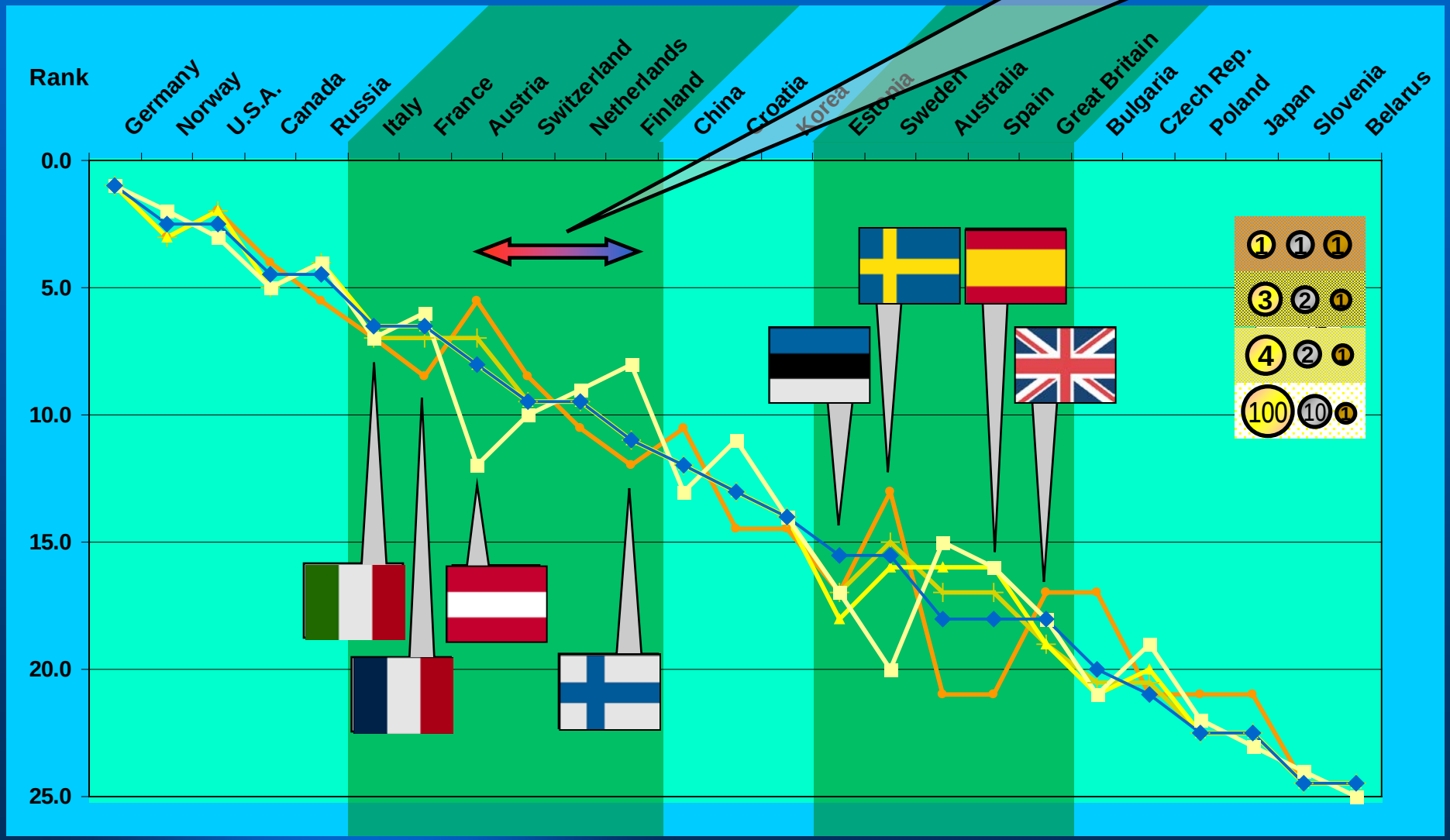
- benign,
- severe, and
- grave

Value of Olympic medals

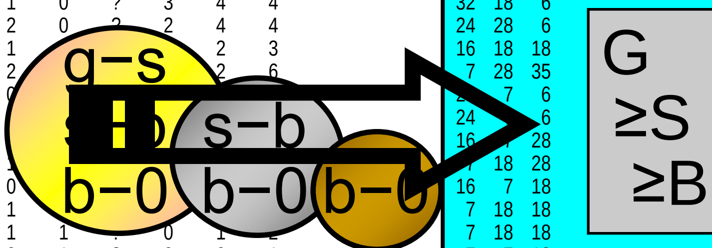
- bronze,
- silver, and
- gold

Salt Lake City 2002

Depending on the weights G:S:B, the rank of **Austria** vs **Finland** can be **6**:**12** or **12**:**8**



Country	g	s	b		G	>=S	>=B	1	1	1	0	0	0	0
Germany	12	16	7	?	12	28	35	50	50	44				
Norway	11	7	6	?	11	18	24	48	46	41				
U.S.A.	10	13	11	?	10	23	34	46	48	50				
Canada	6	3	8	?	6	9	17	43	34	46				
Russia	6	6	4	?	6	12	16	43	44	35				
Italy	4	4	4	?	4	8	12	38	37	35				
France	4	5	2	?	4	9	11	38	41	28				
Austria	2	4	10	?	2	6	16	24	37	48				
Switzerland	3	2	6	?	3	5	11	32	28	41				
Netherland	3	5	0	?	3	8	8	32	41	6				
Finland	4	2	1	?	4	6	7	38	28	18				
China	2	2	4	?	2	4	8	24	28	35				
Croatia	3	1	0	?	3	4	4	32	18	6				
Korea	2	2	0	?	2	4	4	24	28	6				
Estonia	1	1	[illegible]	[illegible]	[illegible]	2	3	16	18	18				
Sweden	0	2	[illegible]	[illegible]	[illegible]	2	6	7	28	35				
Australia	2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	7	6				
Spain	2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	24	[illegible]	6				
Great Brita	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	16	[illegible]	28				
Bulgaria	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	18	28				
Czech Rep	1	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	16	7	18				
Poland	0	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	7	18	18				
Japan	0	1	1	[illegible]	0	[illegible]	[illegible]	7	18	18				
Slovenia	0	0	1	?	0	0	1	7	7	18				
Belarus	0	0	1	?	0	0	1	7	7	18				

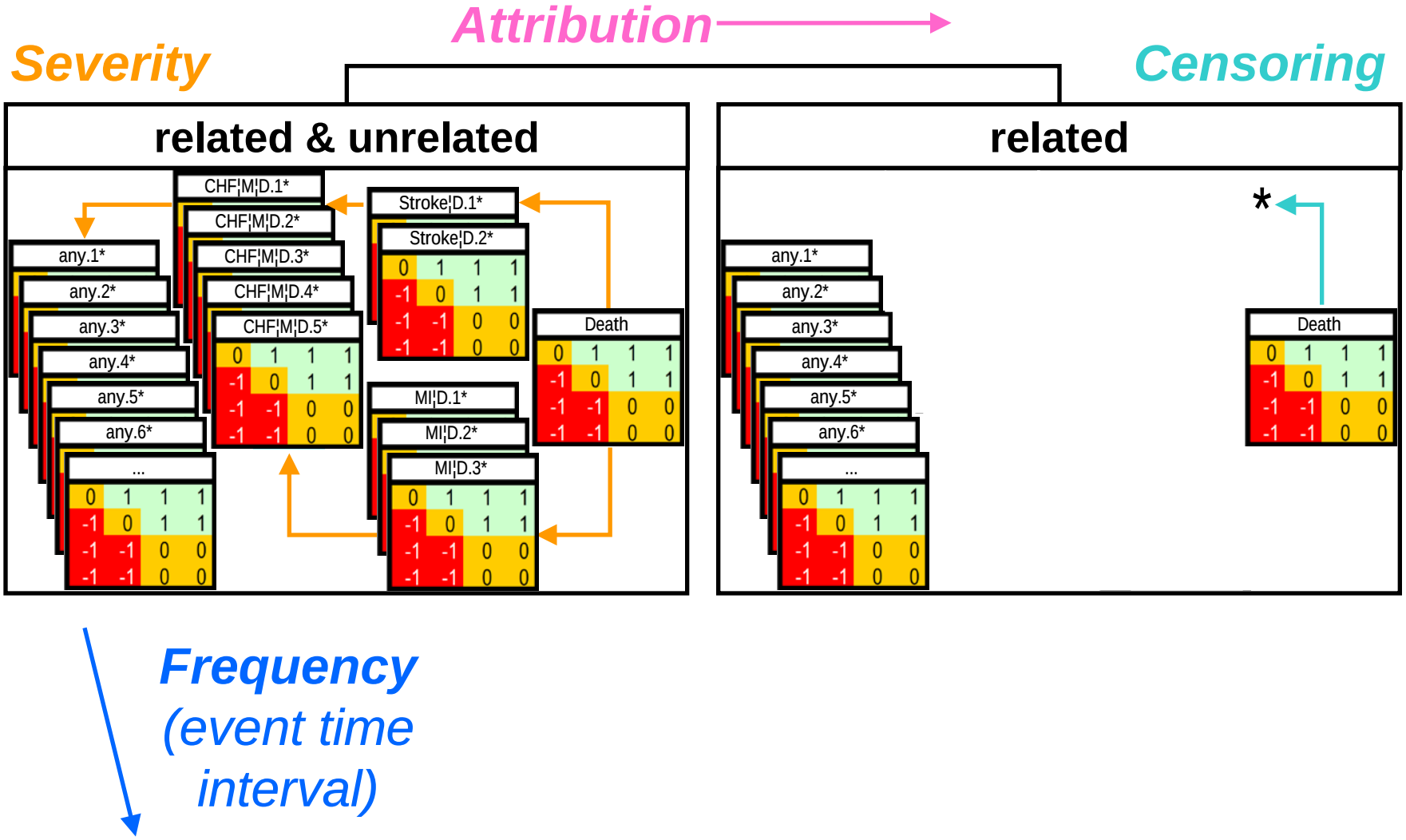


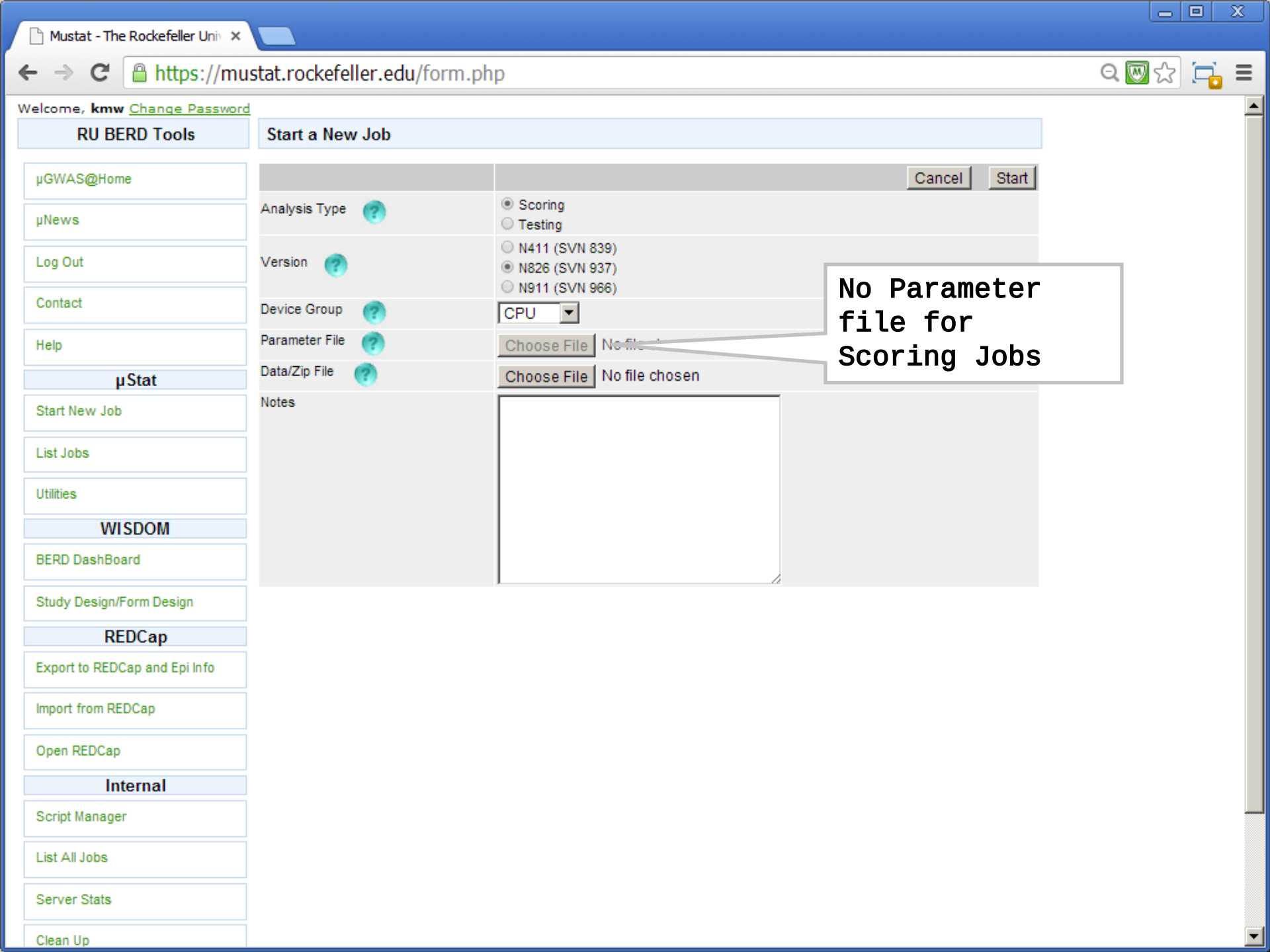
Country	Germany	Norway	U.S.A.	Canada	Russia	Italy	France	Austria	Switzerla	Netherlar	Finland	China	Croatia	Korea	Estonia	Sweden	Australia	Spain	Great Bri	Bulgaria	Czech Re	Poland	Japan	Slovenia	Belarus
Germany	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Norway	-1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
U.S.A.	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Canada	0	0	-1	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Russia	-1	-1	-1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Italy	-1	-1	-1	0	-1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
France	-1	-1	-1	0	-1	0	0	0	0	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1
Austria	0	0	-1	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
Switzerland	-1	-1	-1	-1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Netherland	-1	-1	-1	0	-1	0	-1	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0
Finland	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1
China	-1	-1	-1	-1	-1	-1	0	-1	-1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
Croatia	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	0	0	0	0	1	1	0	0	0	0	0	0	0
Korea	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	1	1	0	0	0	0	0	0	0
Estonia	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	0	0	0	0	0	0	0	0	1	1	1	1	1
Sweden	-1	-1	-1	-1	-1	-1	0	-1	-1	0	0	-1	0	0	0	0	0	0	0	1	0	1	1	1	1
Australia	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	0
Spain	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	0
Great Brita	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	0	0	0	0	0	0	0	0	1	0	0	1	1
Bulgaria	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	0	0	0	-1	0	0	0	0	0	1	1	1	1
Czech Rep	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	0	0	-1	0	0	0	-1	0	0	0	0	1	1
Poland	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	0	0	-1	-1	0	0	0	-1	0	0	0	1	1
Japan	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	0	0	-1	-1	0	0	0	-1	0	0	0	1	1
Slovenia	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	0	0	-1	-1	0	0	-1	-1	-1	-1	-1	0	0
Belarus	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	0	0	-1	-1	0	0	-1	-1	-1	-1	-1	0	0

Country	g	s	b		G	>=S	>=B	0	0	0	0	1	1	1
Germany	12	16	7	?	12	28	35					50	50	50
Norway	11	7	6	?	11	18	24					48	46	46
U.S.A.	10	13	11	?	10	23	34					46	48	48
Canada	6	3	8	?	6	9	17					43	41	44
Russia	6	6	4	?	6	12	16					43	44	41
Italy	4	4	4	?	4	8	12					38	37	38
France	4	5	2	?	4	9	11					38	41	35
Austria	2	4	10	?	2	6	16					24	33	41
Switzerland	3	2	6	?	3	5	11					32	30	35
Netherland	3	5	0	?	3	8	8					32	37	31
Finland	4	2	1	?	4	6	7					38	33	28
China	2	2	4	?	2	4	8					24	26	31
Croatia	3	1	0	?	3	4	4					32	26	23
Korea	2	2	0	?	2	4	4					24	26	23
Estonia	1	1	1	?	1	2	3					16	19	18
Sweden	0	2	4	?	0	2	6					7	19	26
Australia	2	0	0	?	2	2	2					24	19	10
Spain	2	0	0	?	2	2	2					24	19	10
Great Brita	1	0	2	?	1	1	3					16	10	18
Bulgaria	0	1	2	?	0	1	3					7	10	18
Czech Rep	1	0	1	?	1	1	2					16	10	10
Poland	0	1	1	?	0	1	2					7	10	10
Japan	0	1	1	?	0	1	2					7	10	10
Slovenia	0	0	1	?	0	0	1					7	3	3
Belarus	0	0	1	?	0	0	1					7	3	3

Country	Germany	Norway	U.S.A.	Canada	Russia	Italy	France	Austria	Switzerla	Netherlar	Finland	China	Croatia	Korea	Estonia	Sweden	Australia	Spain	Great Bri	Bulgaria	Czech Re	Poland	Japan	Slovenia	Belarus
Germany	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Norway	-1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
U.S.A.	-1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Canada	-1	-1	-1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Russia	-1	-1	-1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Italy	-1	-1	-1	-1	-1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
France	-1	-1	-1	-1	-1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Austria	-1	-1	-1	-1	-1	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
Switzerland	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Netherland	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Finland	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
China	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
Croatia	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	0	1	1	0	1	1	1	1	1	1	1	1	1
Korea	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	1	0	1	1	1	1	1	1	1	1	1
Estonia	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	1	1	1	1	1	1	1
Sweden	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0	1	0	1	1	1	1
Australia	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	1	1	1	1	1
Spain	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	1	1	1	1	1
Great Brita	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	1	1	1	1	1	1
Bulgaria	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	0	0	1	1	1	1
Czech Rep	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	0	1	1	1	1
Poland	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	1	1
Japan	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	1	1
Slovenia	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0
Belarus	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0

µPhene Data Structure




Mustat - The Rockefeller Uni
https://mustat.rockefeller.edu/form.php
Welcome, kmw Change Password
RU BERD Tools
μGWAS@Home
μNews
Log Out
Contact
Help
μStat
Start New Job
List Jobs
Utilities
WISDOM
BERD DashBoard
Study Design/Form Design
REDCap
Export to REDCap and Epi Info
Import from REDCap
Open REDCap
Internal
Script Manager
List All Jobs
Server Stats
Clean Up
Start a New Job
Cancel
Start
Analysis Type
Scoring
Testing
Version
N411 (SVN 839)
N826 (SVN 937)
N911 (SVN 966)
Device Group
CPU
Parameter File
Choose File
Data/Zip File
Choose File
No file chosen
Notes
No Parameter file for Scoring Jobs

Data Preparation: Cum. Variables

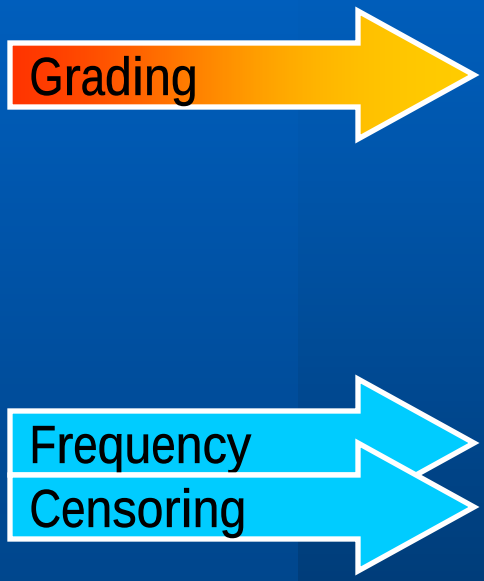
```
days until        (                                    death)
```

```
days until 1st of (any                  stroke |       death)
days until 2nd of (any                  stroke |       death)

days until 1st of (any                      MI | death)
days until 2nd of (any                      MI | death)
days until 3rd of (any                      MI | death)
```

```
days until 1st of (any          CHF | stroke | MI | death)
days until 2nd of (any          CHF | stroke | MI | death)
...
days until nth of (any          CHF | stroke | MI | death)
```

```
days until 1st of (any other | CHF | stroke | MI | death)
days until 2nd of (any other | CHF | stroke | MI | death)
...
days until nth of (any other | CHF | stroke | MI | death)
```

µPhene Data Structure



Hdr2	Death	Death/Stroke	Death/Myocardial Infarction	Death/Stroke/MI/ CHF	Any Complications
Pht1	1				
Pht2	1	2		3	4
Pht3		1	2		
Pht4	1	1 2	1 2 3	1 n	1 n
Ivl		-1 1 -2 2	-3 3 -4 4 -5 5	... ...	... ...
Pol	1	1 1 1 1	1 1 1 1 1 1	1 1 1 1 1	1 1 1 1 1
PID					
Dat1					
Dat2					
Dat3					
Dat4					
...					

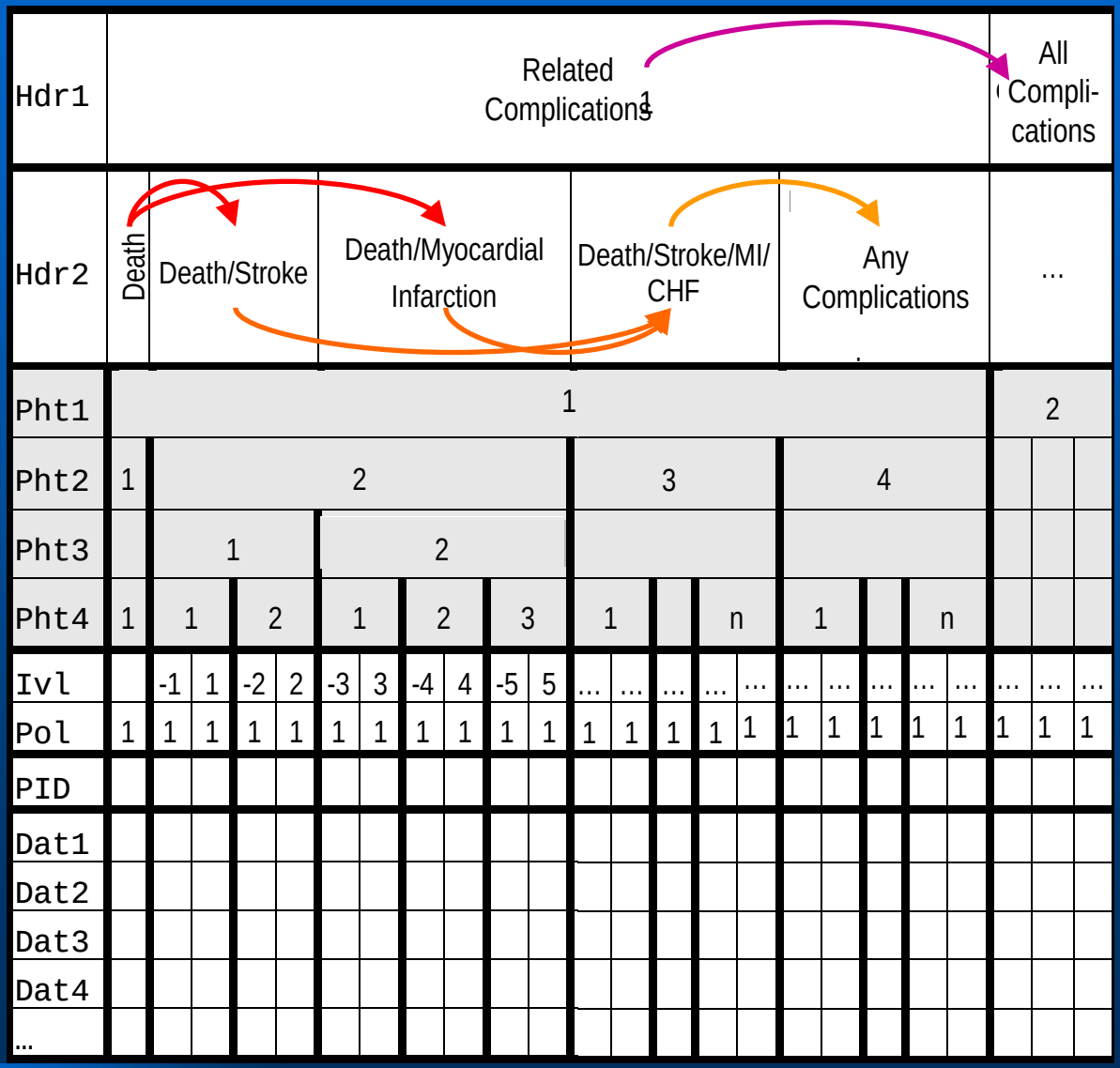
µPhene Data Structure

Attribution

Grading

Frequency

Censoring



Combining Endpoint Evidence

With u-scores for multivariate ordinal data (μ-Scores), one can objectively score

- Multiple
- Graded
- Repeated
- Censored

Events.

2011: μ-scores accepted by FDA for Phase III

2009: What are the remaining problems?

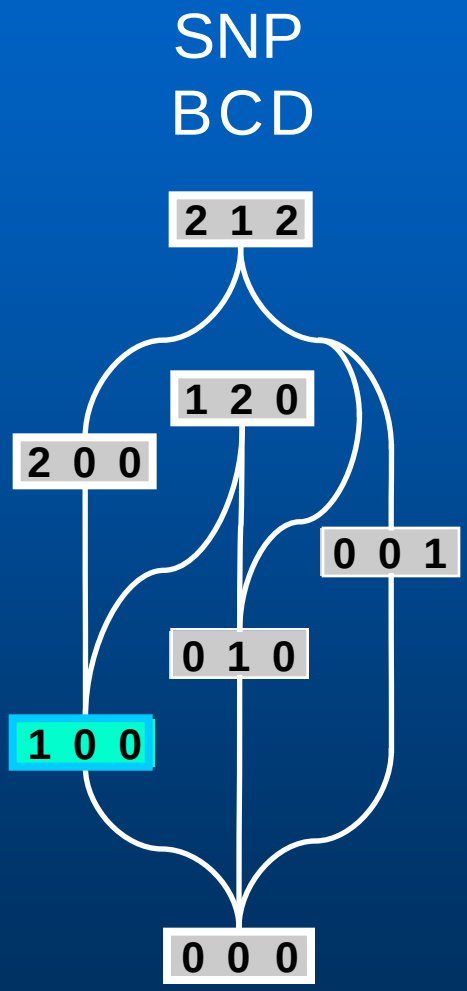
- Sufficiently detailed data remains elusive as long as
- Regulatory agencies are satisfied with crude the measures
  - Time to first event
  - Overall survival

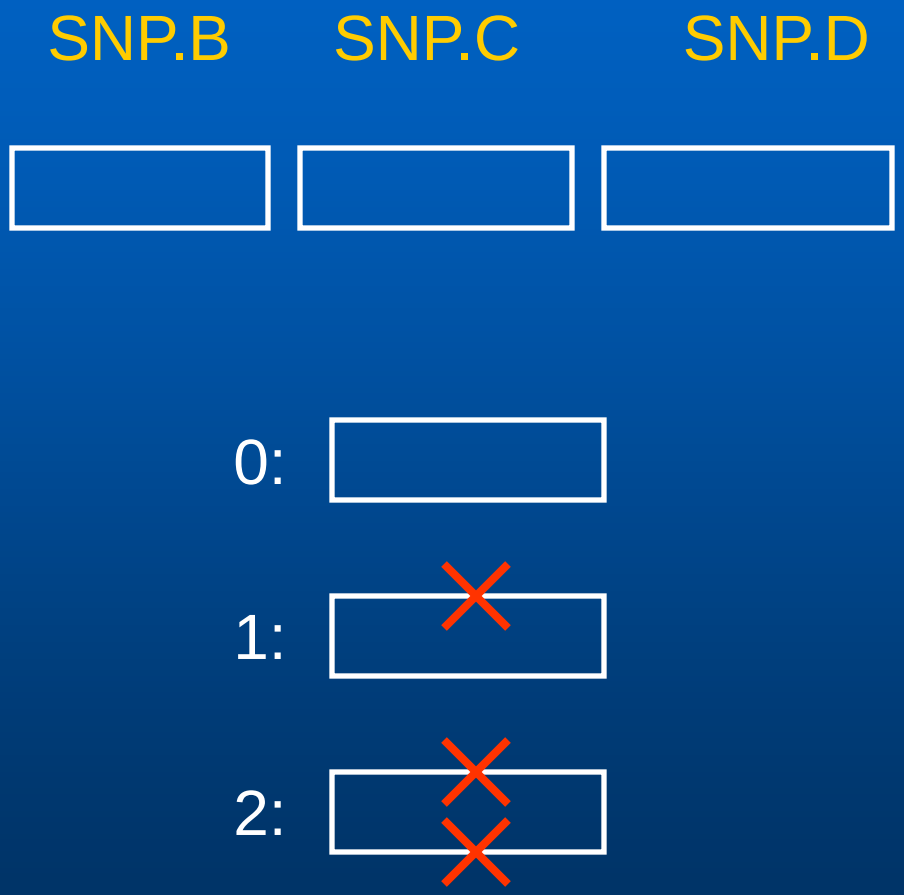
?

U-Statistics for Multivariate Data:

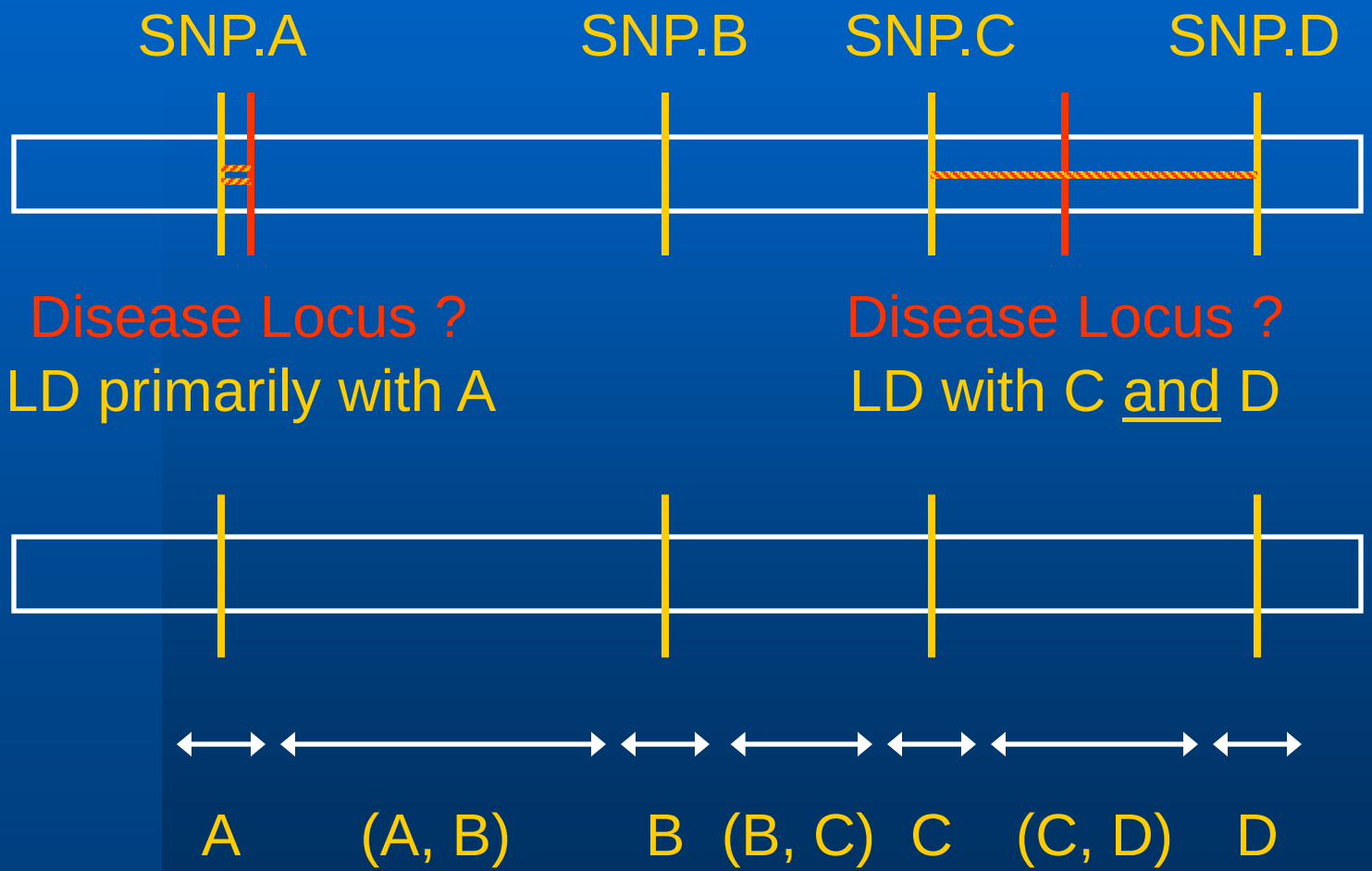
GWAS

More than one SNP

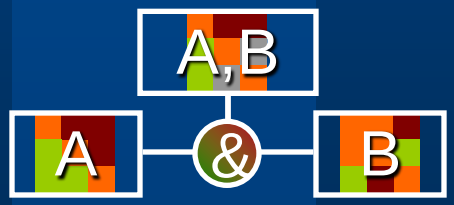




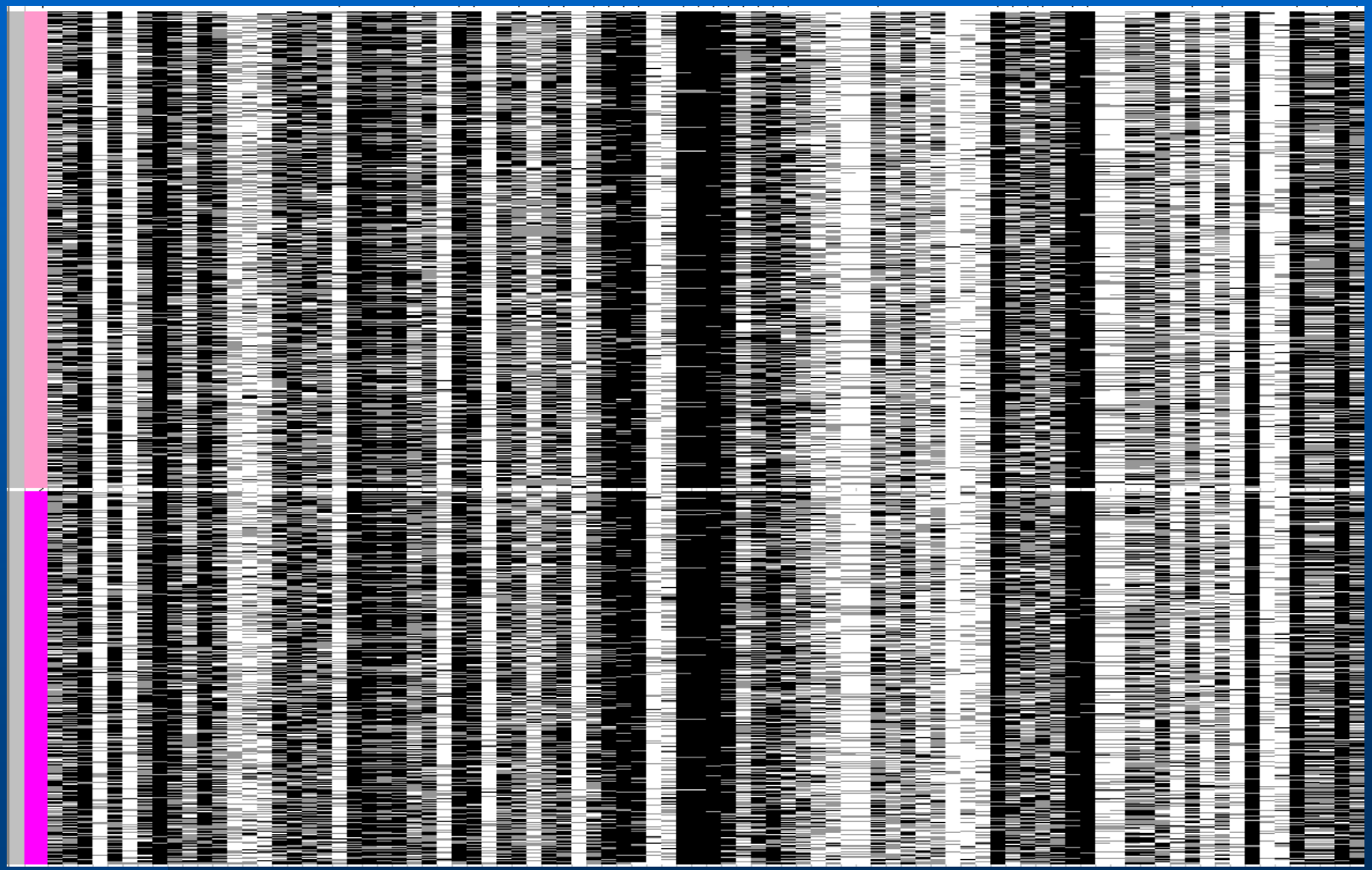
Linkage Disequilibrium (LD)



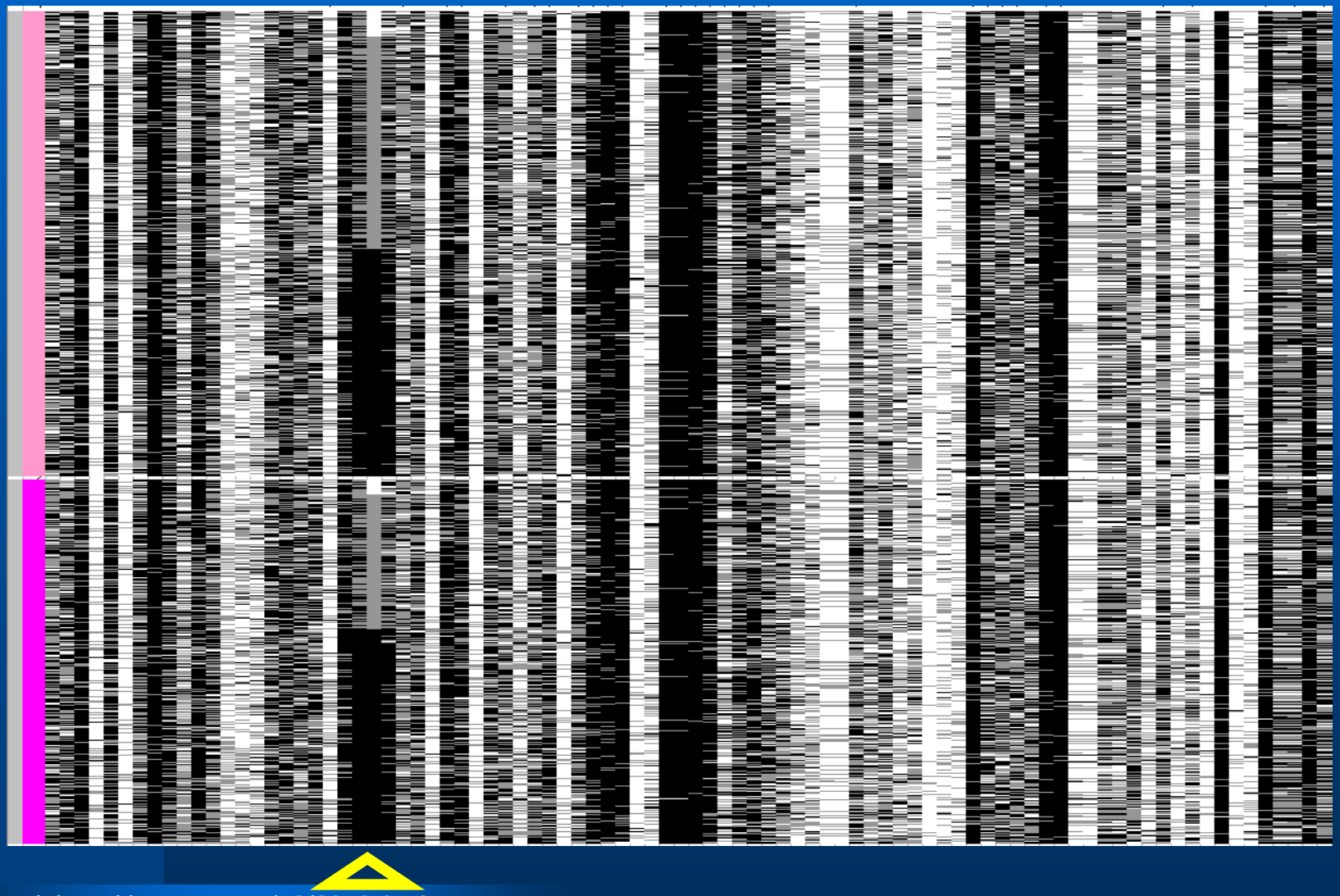
Analysis of Genetic Intervals



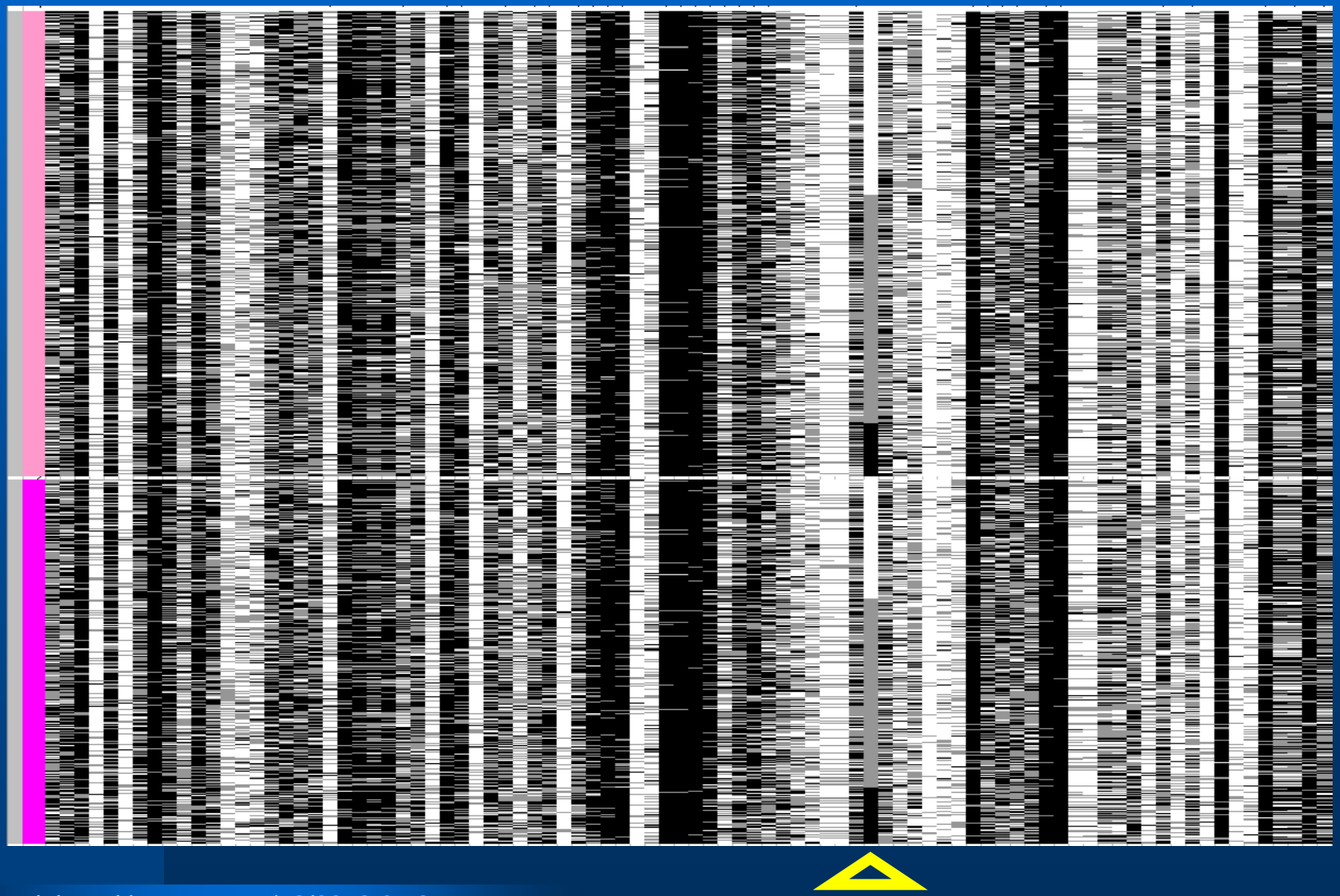
SNP Data sorted by ID



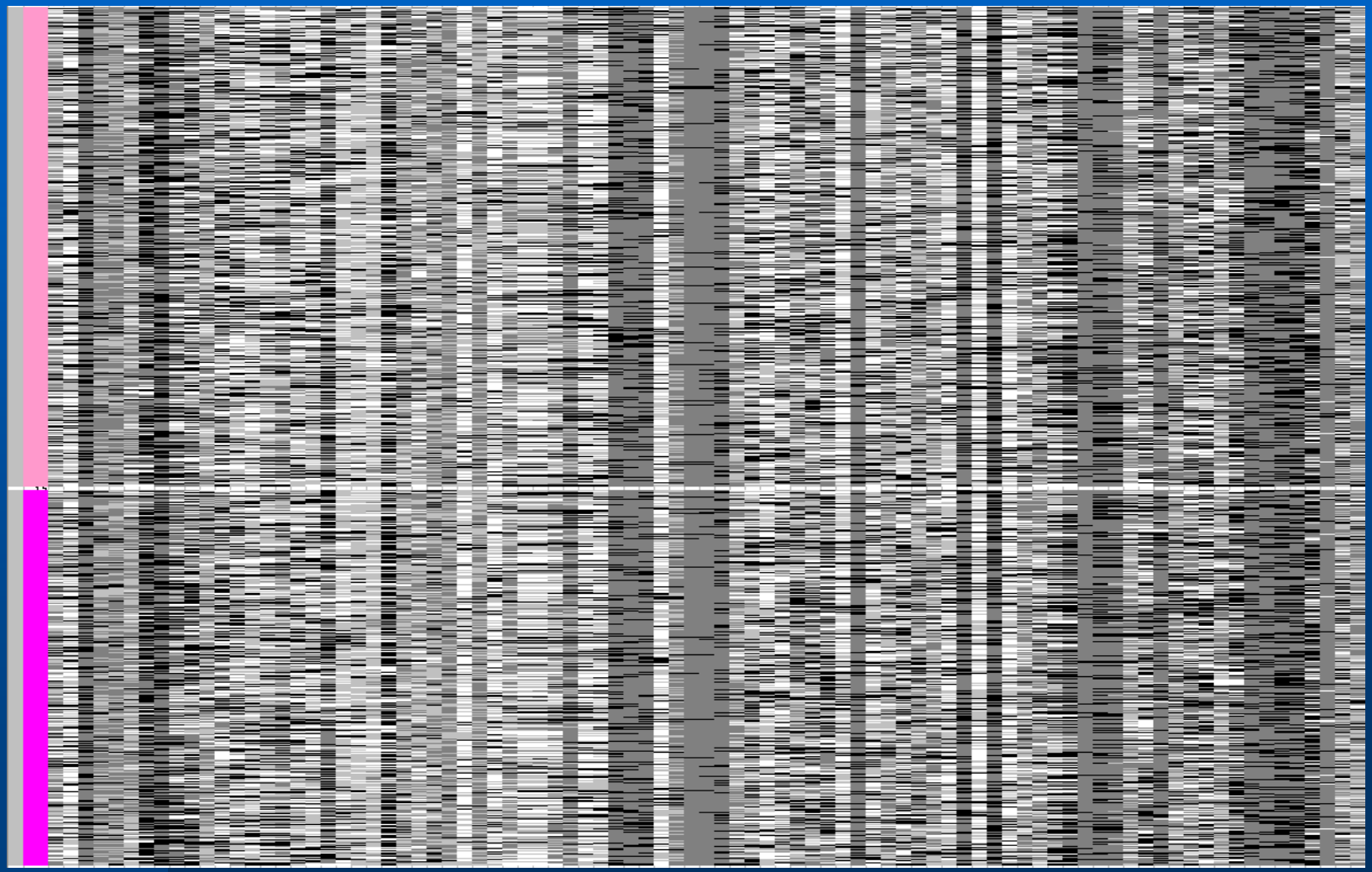
SNP Data Sorted by 1st signif SNP



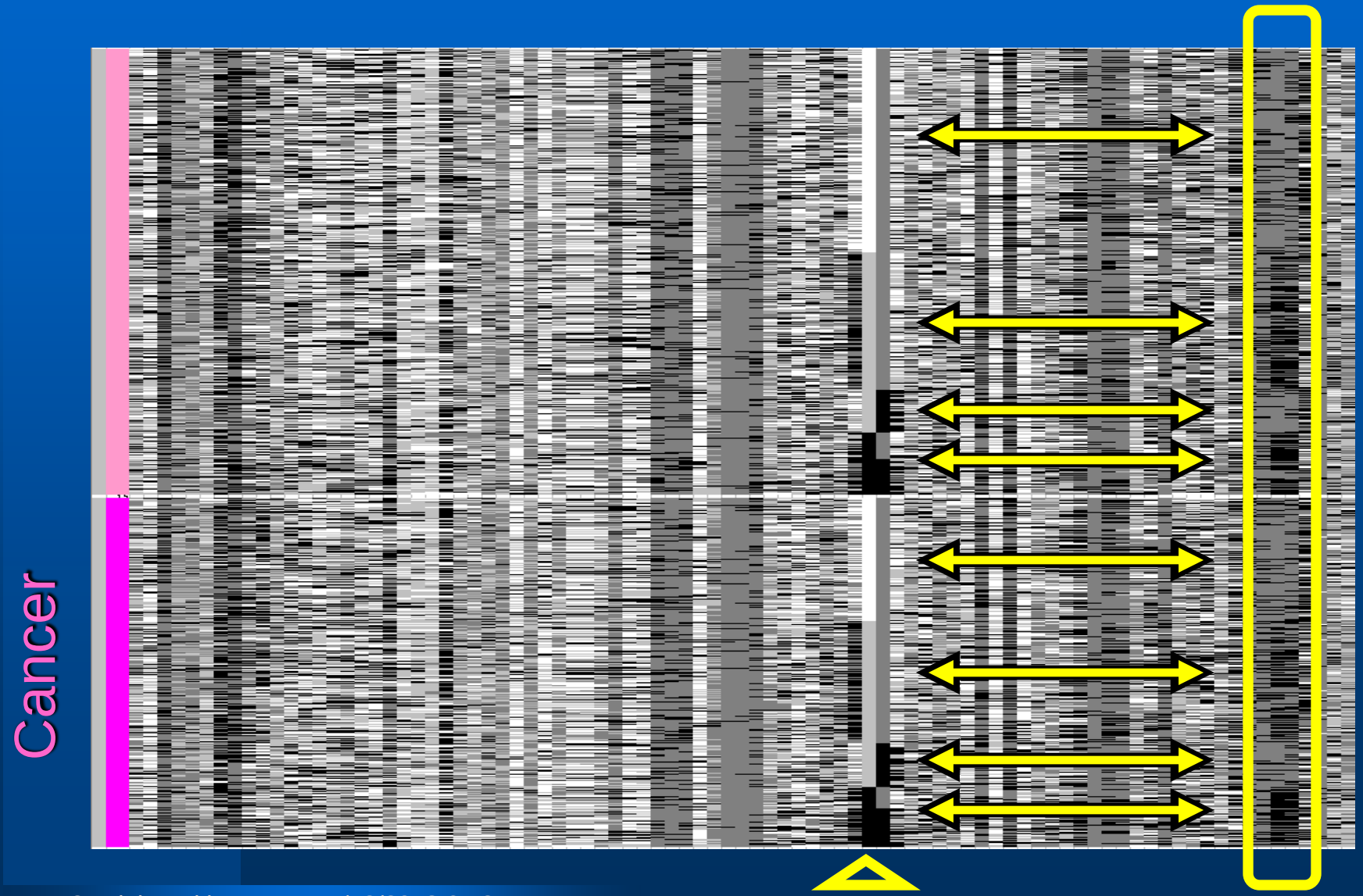
SNP Data Sorted by 2nd signif SNP



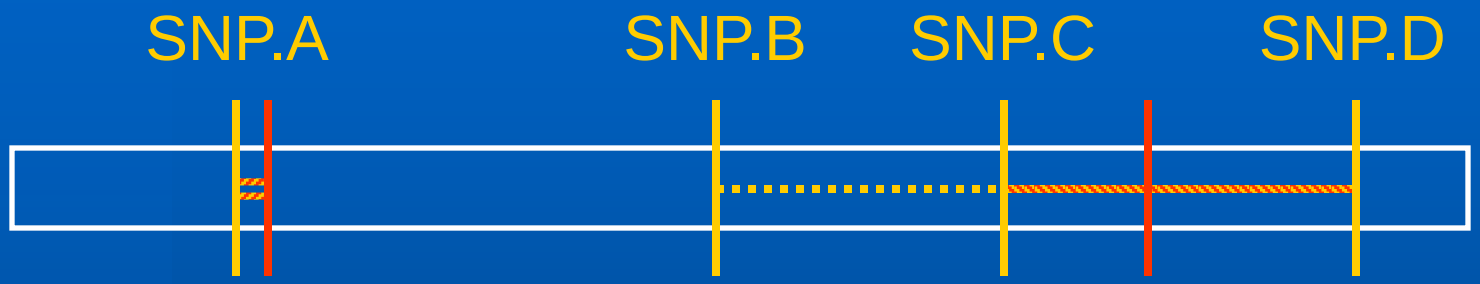
Interval Scores Sorted by ID



Interval Scores by 1st signif Interval

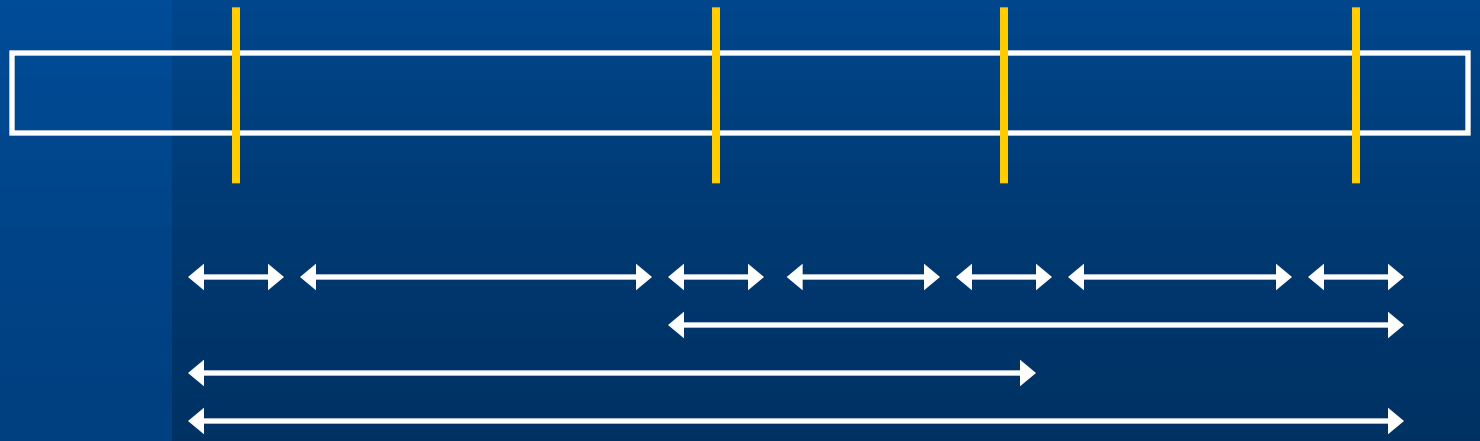


Extended Disequilibrium

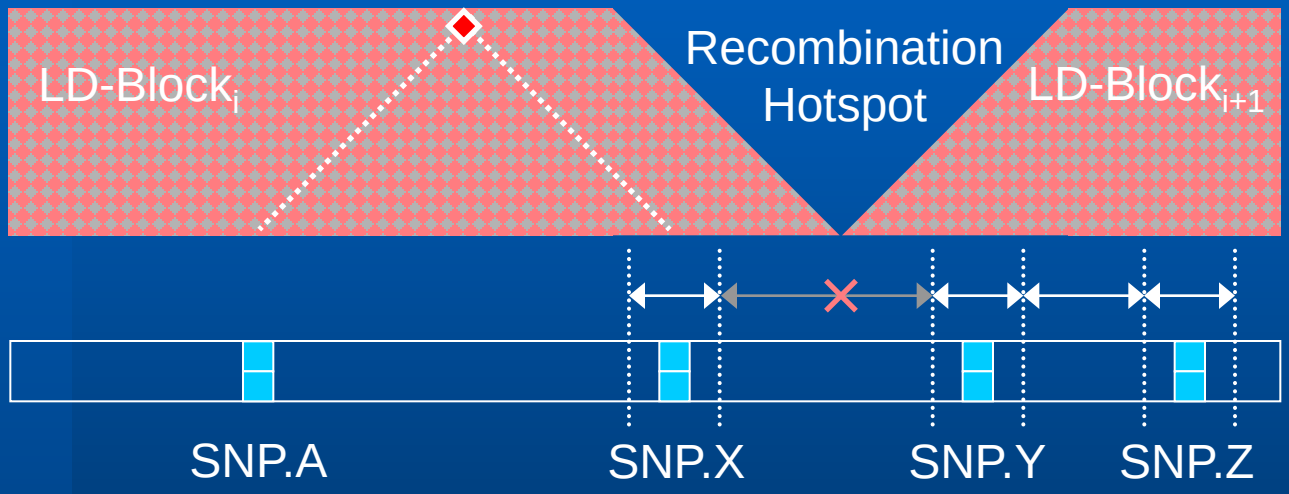


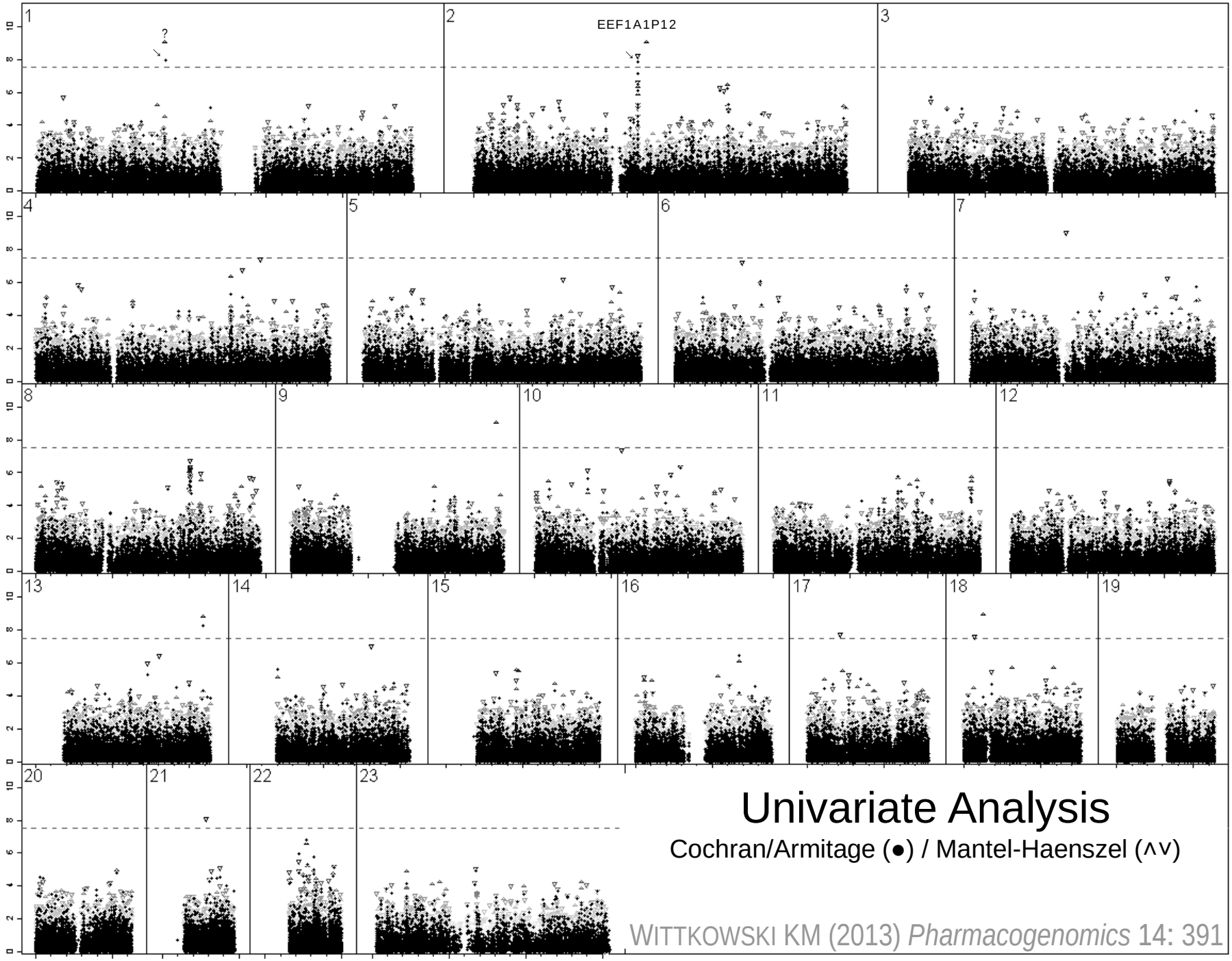
Proximal Disease Locus
LD primarily with A

Intermediate Disease Locus
LD with C and D (and B)

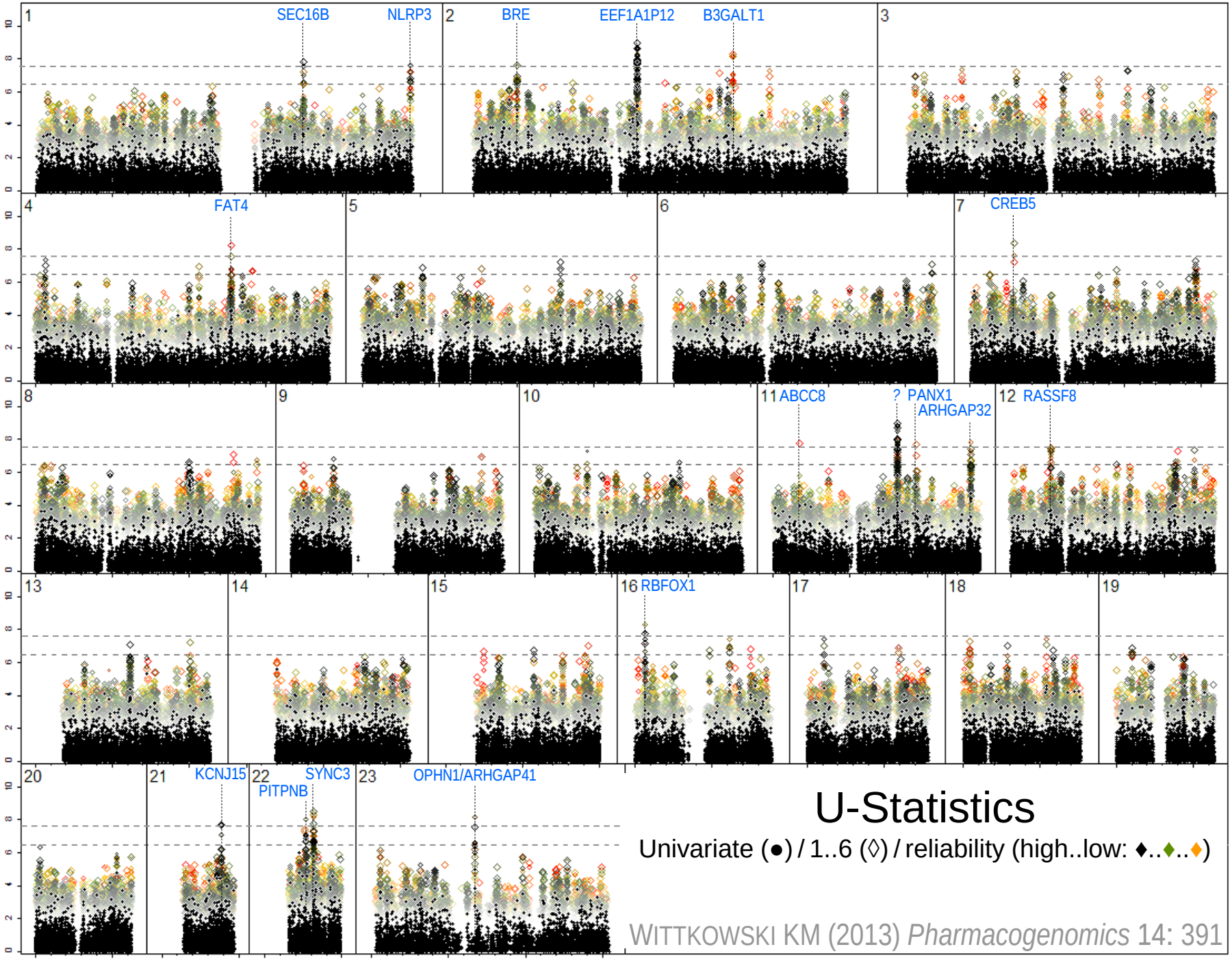


Linkage Disequilibrium (LD)

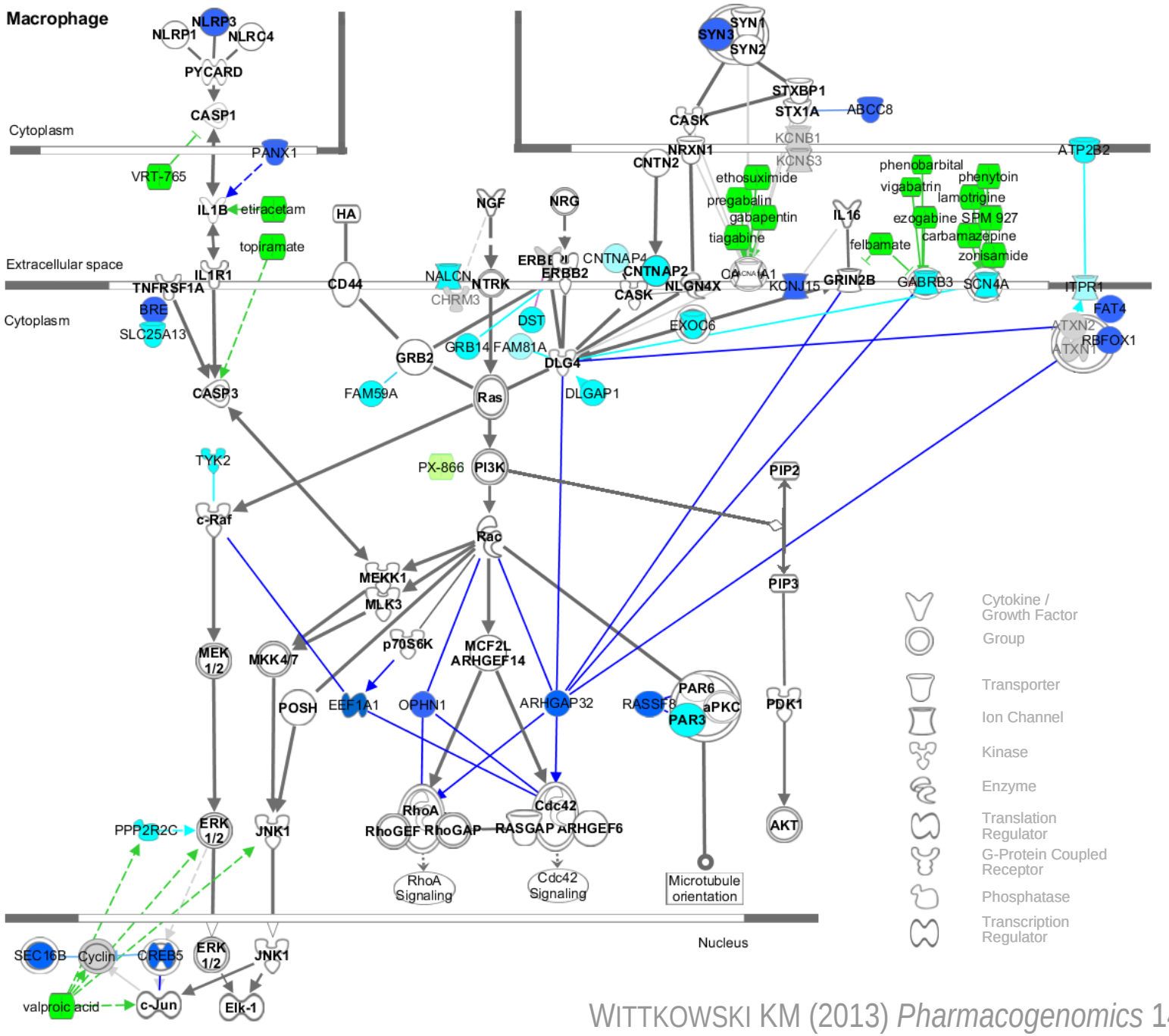




WITTKOWSKI KM (2013) *Pharmacogenomics* 14: 391



WITTKOWSKI KM (2013) *Pharmacogenomics* 14: 391

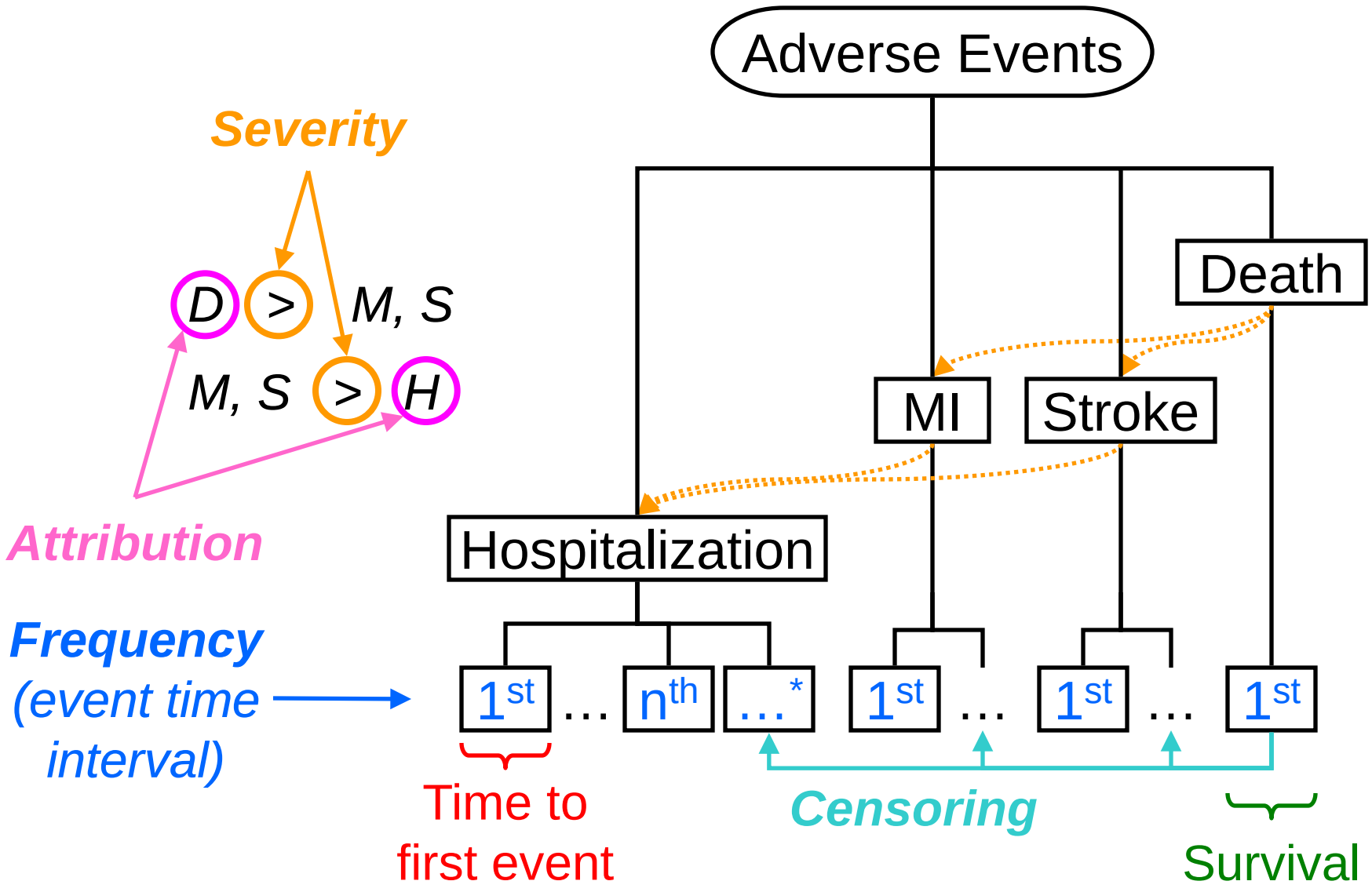

Macrophage
NLRP3
NLRP1
NLRC4
PYCARD
CASP1
Cytoplasm
PANX1
VRT-765
IL1B
etiracetam
topiramate
Extracellular space
IL1R1
TNFRSF1A
BRE
SLC25A13
Cytoplasm
CASP3
TYK2
c-Raf
MEK 1/2
MKK4/7
MEKK1
MLK3
p70S6K
POSH
EEF1A1
OPHN1
MCF2L ARHGEF14
ARHGAP32
RASSF8
PAR6
PAR3
aPKC
PDK1
PIP2
PIP3
AKT
PX-866
PI3K
Rac
Ras
DLG4
DLGAP1
GRB2
FAM59A
GRB14
FAM81A
DST
EXOC6
HA
CD44
NALCN
CHRM3
NTRK
NGF
NRG
ERBB2
CNTNAP4
CNTNAP2
NLGN4X
CASK
KCNJ15
GRIN2B
GABRB3
SCN4A
ITPR1
FAT4
ATXN2
ATXN1
RBFOX1
SYN1
SYN3
SYN2
STXBP1
STX1A
ABCC8
KCNB1
KCNS3
CNTN2
NRXN1
ATP2B2
IL16
ethosuximide
pregabalin
gabapentin
tiagabine
phenobarbital
vigabatrin
phenytoin
lamotrigine
ezogabine
SPM 927
carbamazepine
zonisamide
felbamate
PPP2R2C
ERK 1/2
JNK1
RhoA
RhoGEF RhoGAP
Cdc42
RASGAP ARHGEF6
RhoA Signaling
Cdc42 Signaling
Microtubule orientation
Nucleus
SEC16B
Cyclin
CREB5
ERK 1/2
JNK1
valproic acid
c-Jun
Elk-1
Cytokine / Growth Factor
Group
Transporter
Ion Channel
Kinase
Enzyme
Translation Regulator
G-Protein Coupled Receptor
Phosphatase
Transcription Regulator

Putting it all together:

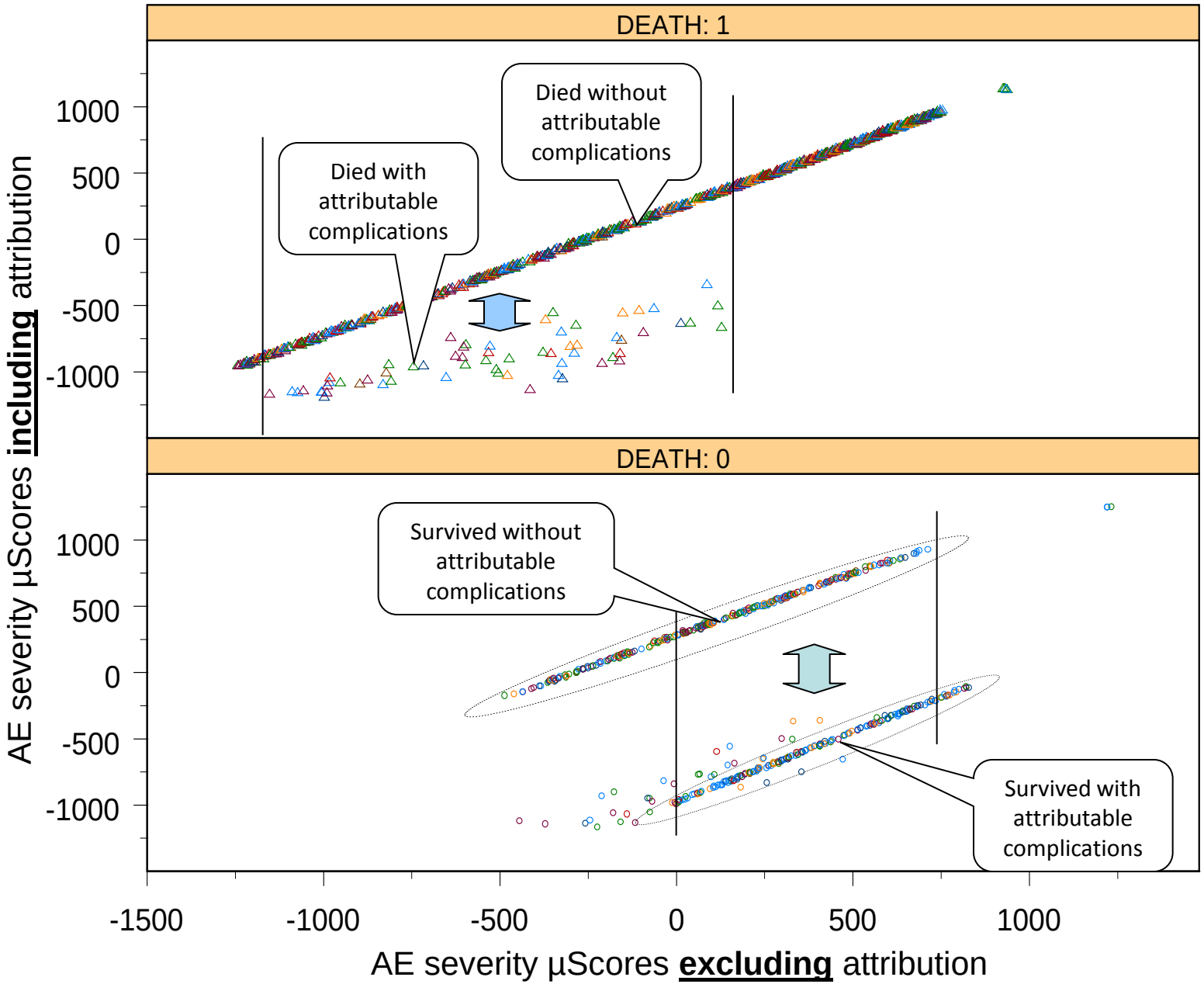
“Gold mining” Phase III Data

- Non-Response
- Adverse Events

µPhene Combines Complex Outcomes

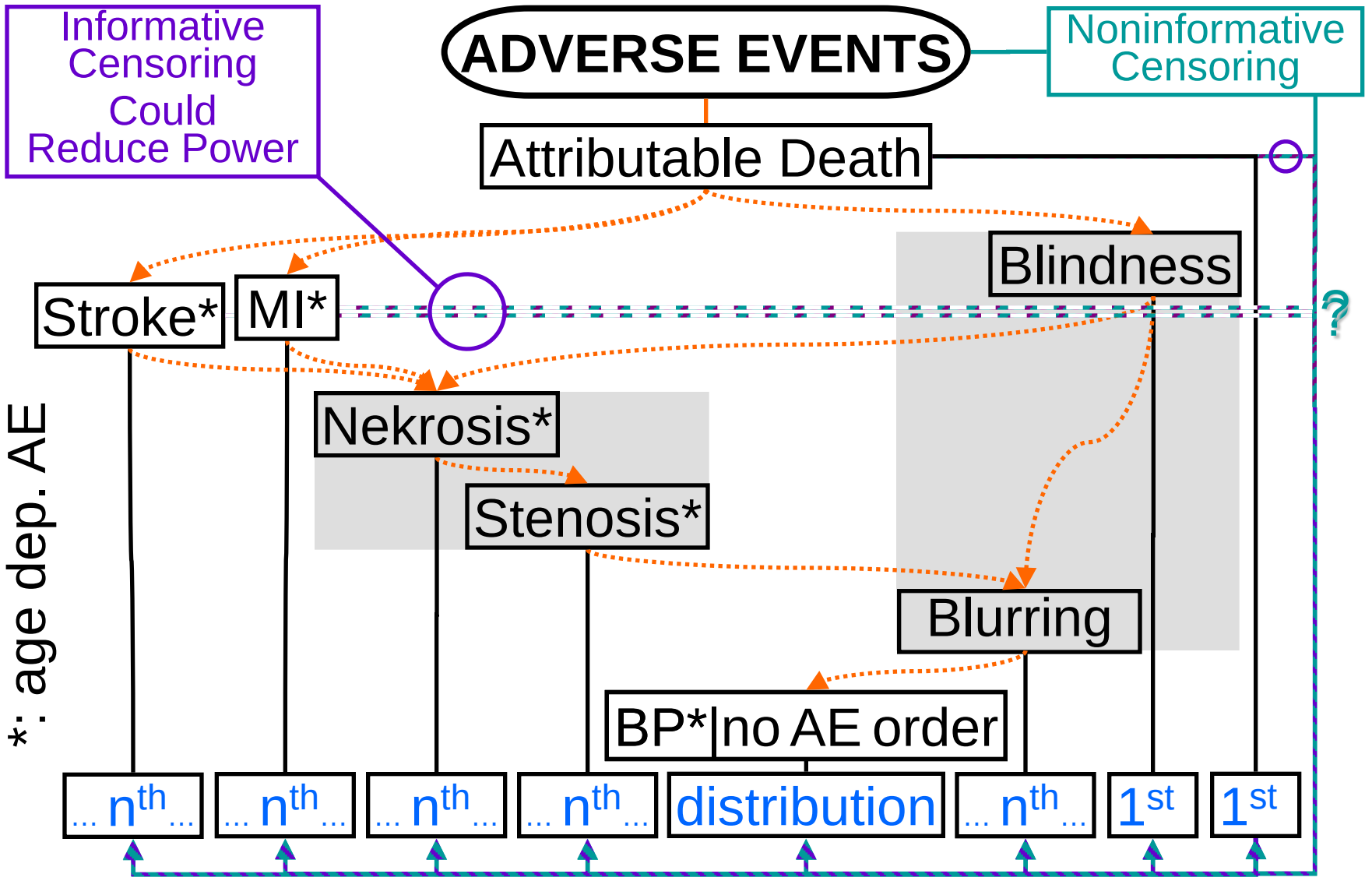


“Gold-mining” Phase III Data



WITTKOWSKI KM (1988) *JASA*

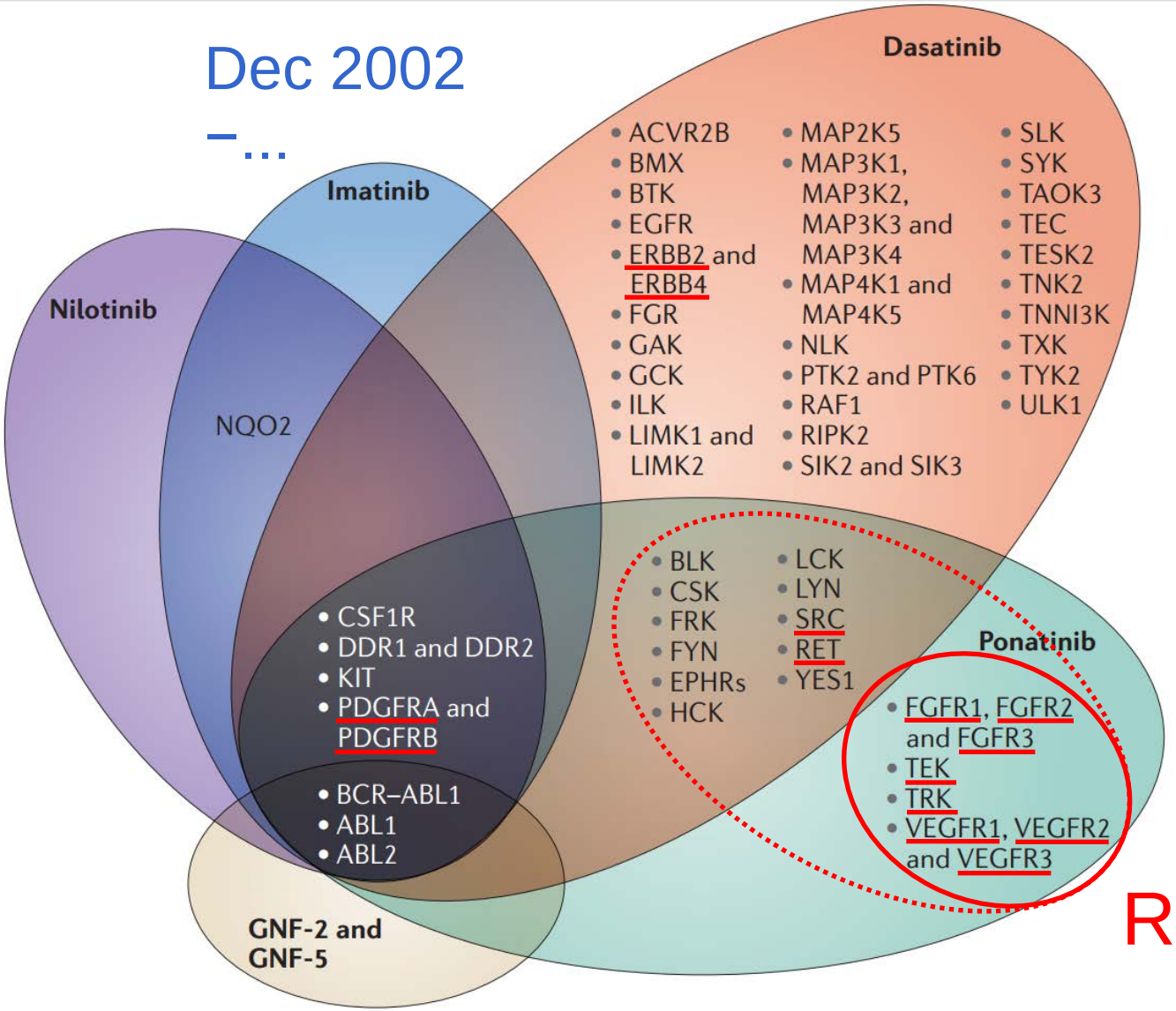
µPhene Combines Complex Outcomes



Targets of BCR-ABL Inhibitors

Dec 2002 –…

Oct 2010 –…

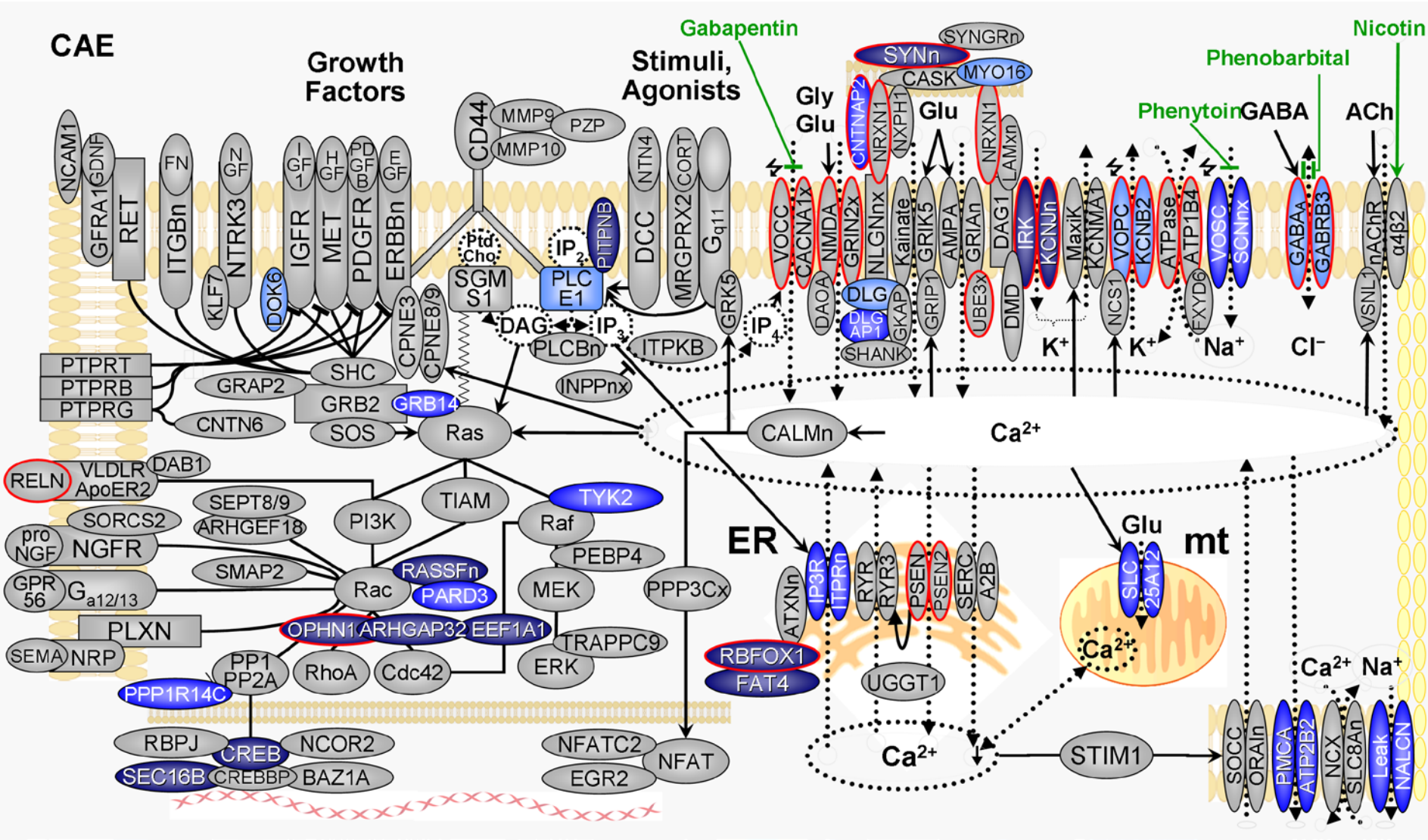


Dec 2012 – Nov 2013

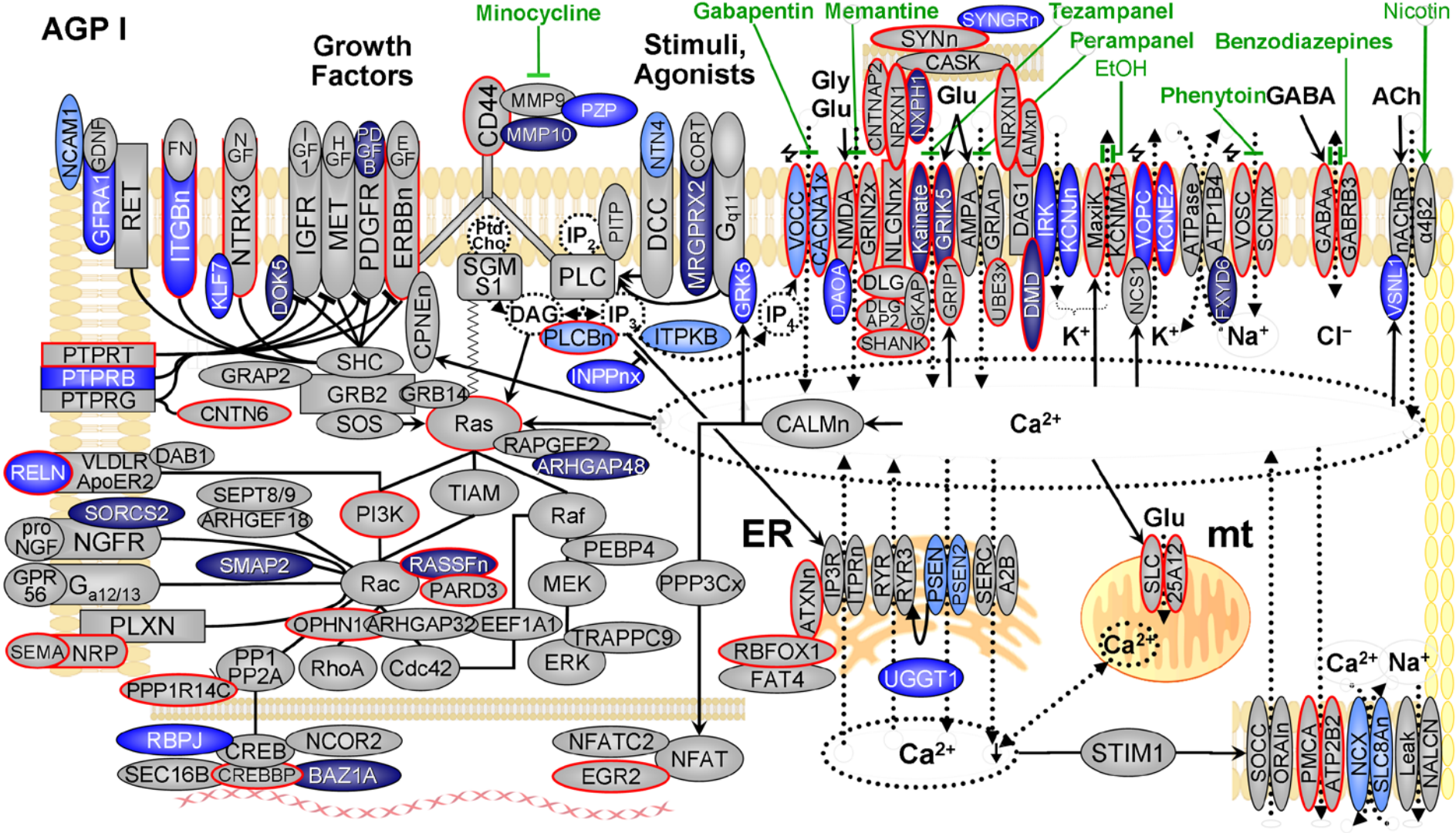
RTK/Ras ?

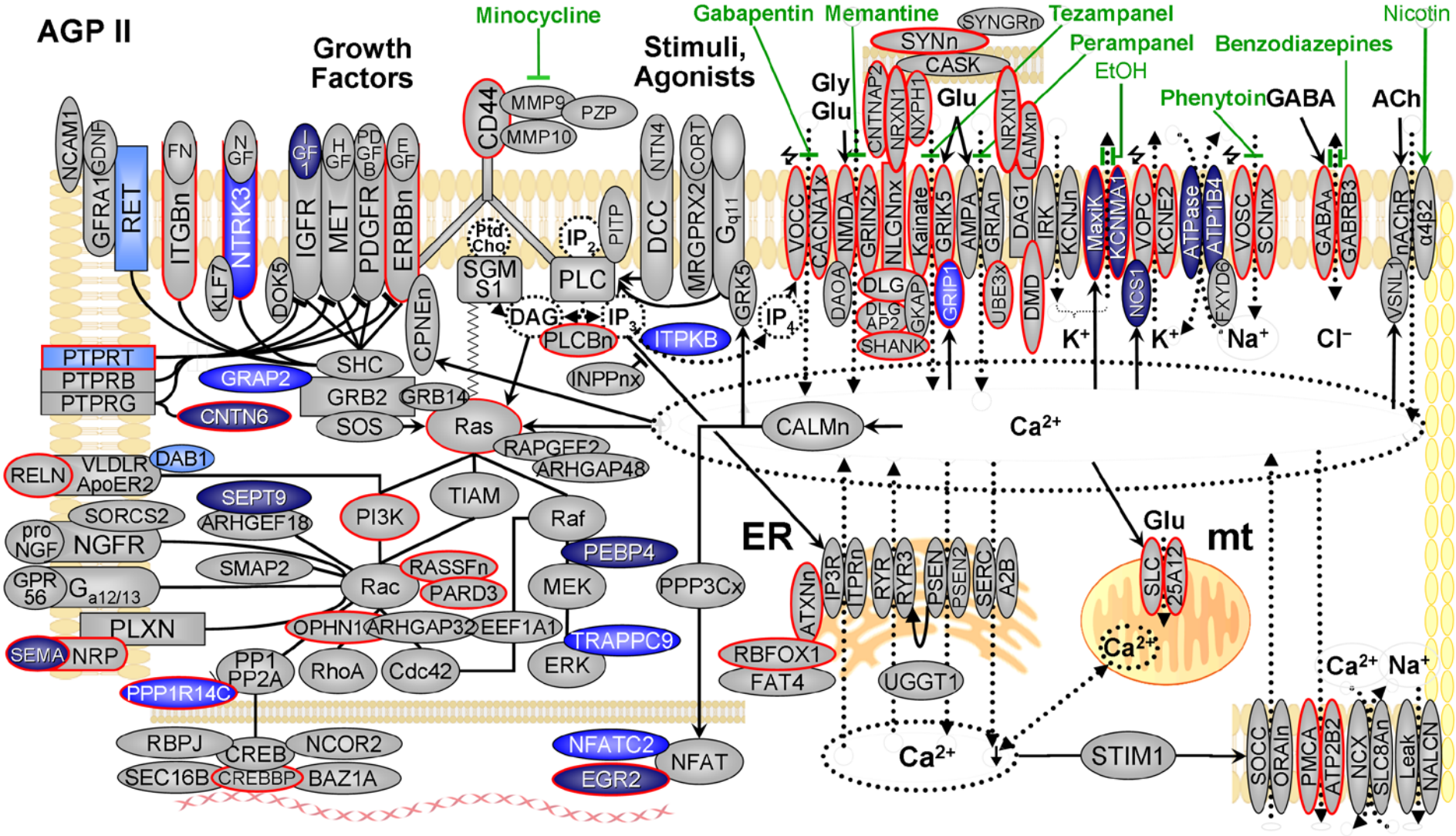
Results:

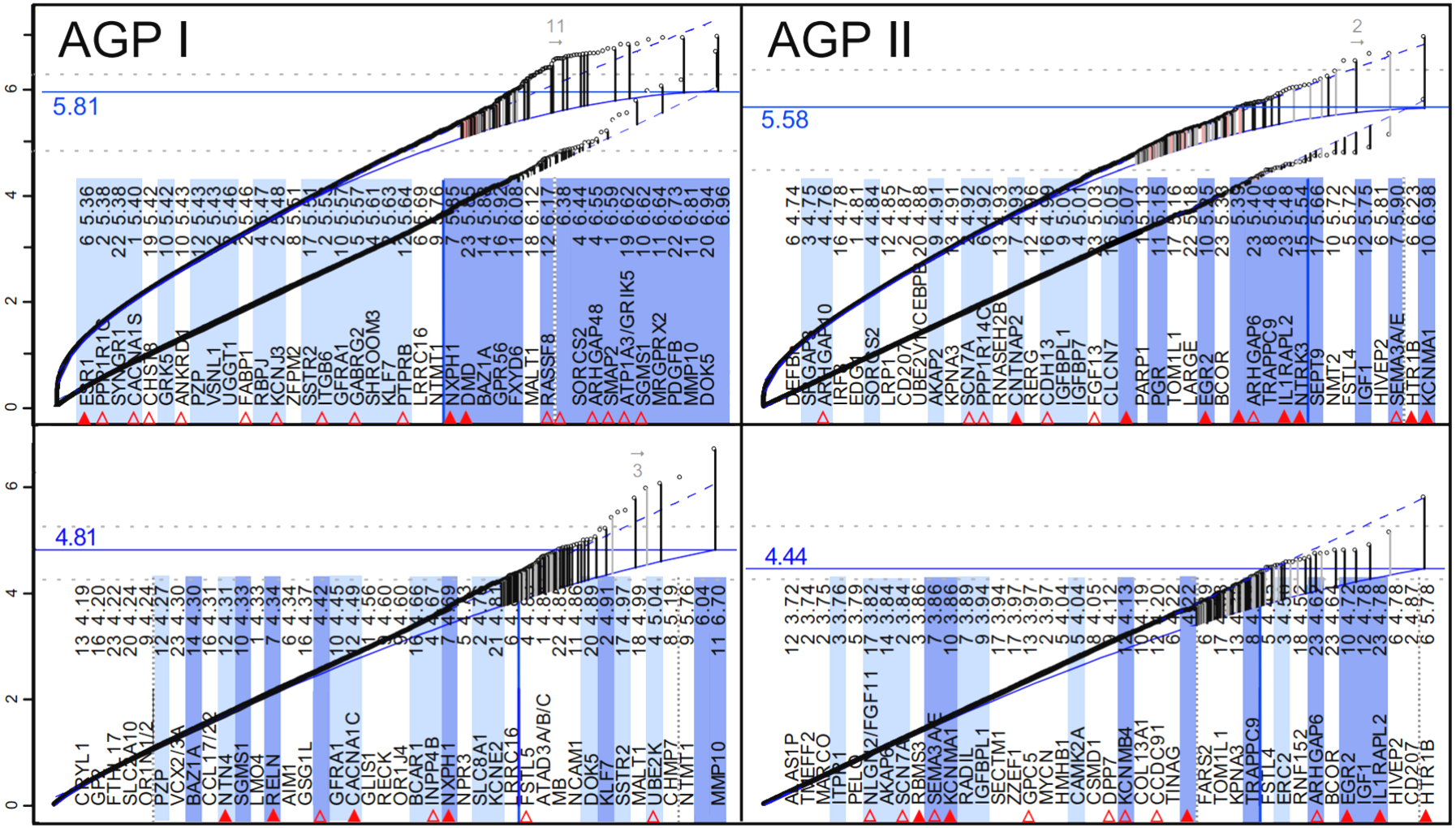
Ras pathway (Autism GWAS)

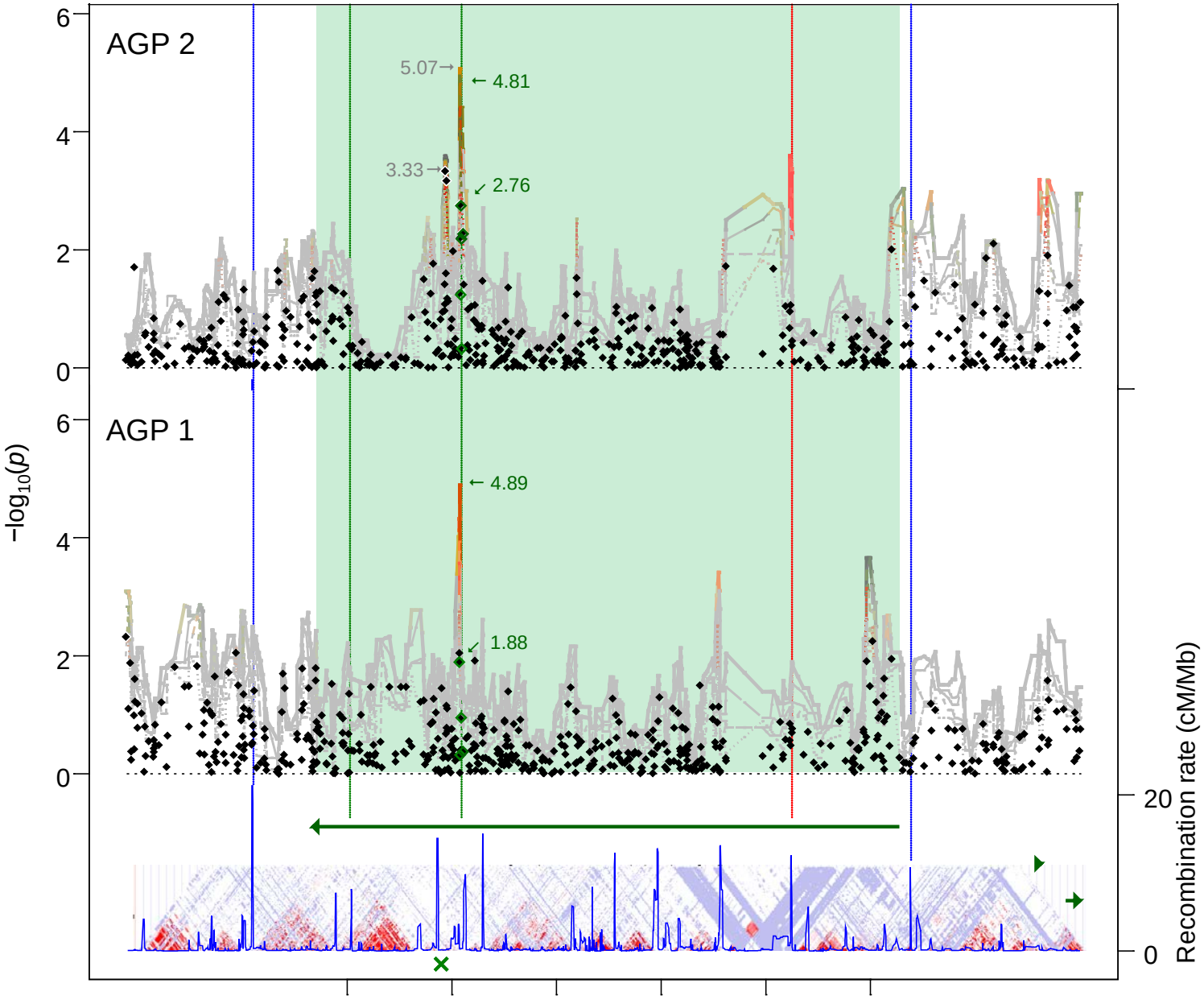


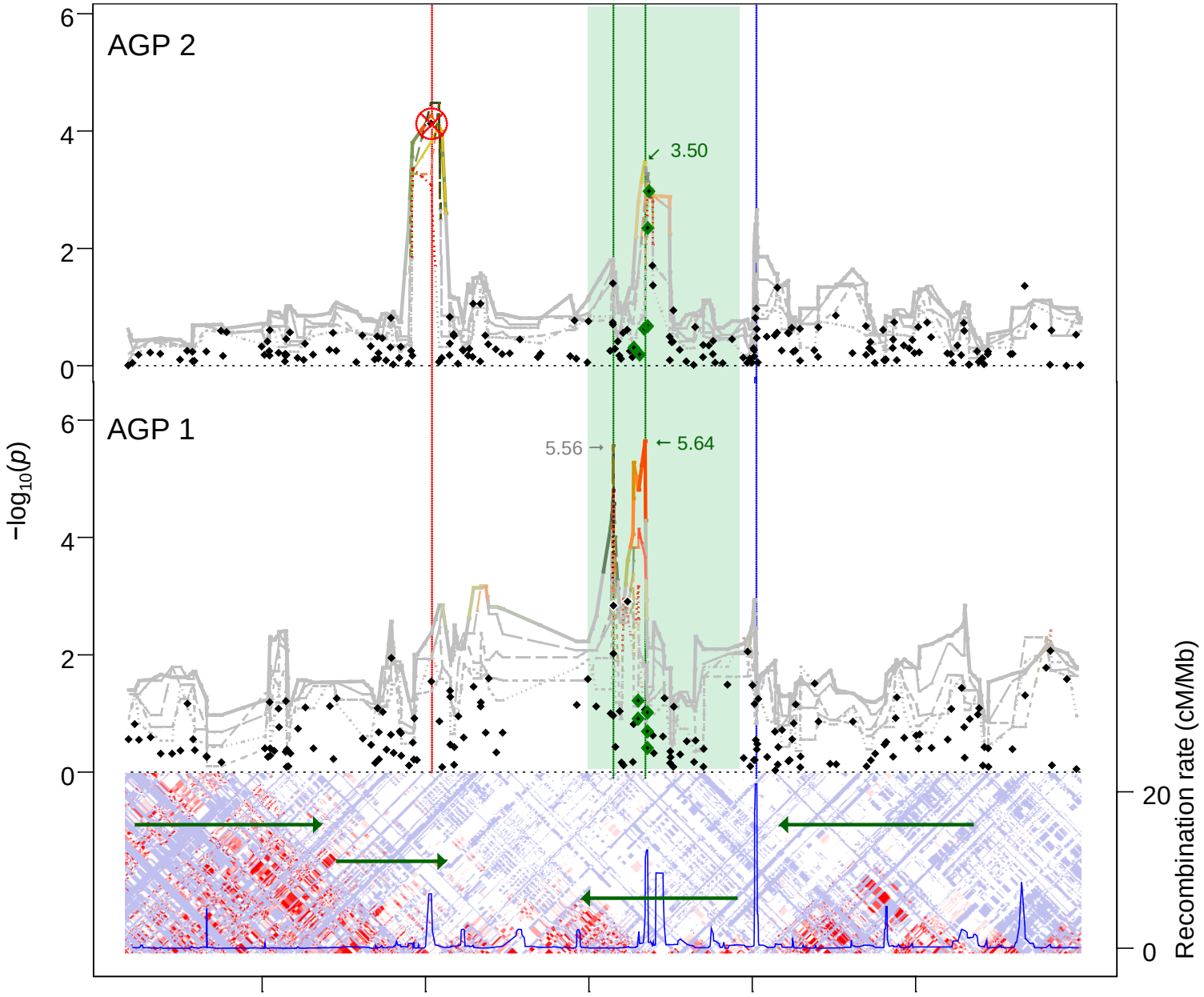
WITTKOWSKI KM (2013) *Pharmacogenomics* 14: 391


AGP I
Growth Factors
Stimuli, Agonists
Minocycline
Gabapentin
Memantine
Tezampanel
Perampanel
Benzodiazepines
Nicotin
EtOH
Phenytoin
GABA
ACh
Gly
Glu
SYNGRn
SYNn
CASK
NCAM1
GDNF
GFRA1
RET
FN
ITGBn
NGF
NTRK3
IGF1
IGFR
HGF
MET
PDGFB
PDGFR
EGF
ERBBn
DOK5
KLF7
CD44
MMP9
MMP10
PZP
NTN4
DCC
CORT
MRGPRX2
Gq11
GRK5
CPNEn
Ptd Cho
SGM S1
PLC
PITP
IP2
DAG
IP3
PLCBn
ITPKB
INPPnx
IP4
PTPRT
PTPRB
PTPRG
SHC
GRAP2
GRB2
GRB14
SOS
Ras
CNTN6
RAPGEF2
ARHGAP48
RELN
VLDLR
ApoER2
DAB1
SEPT8/9
ARHGEF18
PI3K
TIAM
Raf
SORCS2
pro NGF
NGFR
SMAP2
RASSFn
PARD3
Rac
PEBP4
MEK
PPP3Cx
GPR 56
G a12/13
PLXN
SEMA
NRP
OPHN1
ARHGAP32
EEF1A1
TRAPPC9
ERK
PP1
PP2A
RhoA
Cdc42
PPP1R14C
RBPJ
CREB
NCOR2
SEC16B
CREBBP
BAZ1A
NFATC2
NFAT
EGR2
CALMn
Ca2+
ER
ATXNn
IP3R
ITPRn
RYR
RYR3
PSEN
PSEN2
SERC
A2B
RBFOX1
FAT4
UGGT1
Glu
mt
SLC
25A12
Ca2+
Ca2+
Na+
STIM1
SOCC
ORAIn
PMCA
ATP2B2
NCX
SLC8An
Leak
NALCN
CNTNAP2
NRXN1
NXPH1
NRXN1
LAMxn
VOCC
CACNA1x
NMDA
GRIN2x
NLGNnx
Kainate
GRIK5
AMPA
GRIAn
DAG1
IRK
KCNJn
MaxiK
KCNMA1
VOPc
KCNE2
ATPase
ATP1B4
FXYD6
VOSC
SCNnx
GABAA
GABRB3
VSNL1
nAChR
α4β2
DAOA
DLG
DLG AP2
GKAP
SHANK
GRIP1
UBE3x
DMD
NCS1
K+
K+
Na+
Cl−


AGP II
Growth Factors
Stimuli, Agonists
Minocycline
Gabapentin
Memantine
SYNGRn
Tezampanel
Nicotin
SYNn
Perampanel
Benzodiazepines
CASK
EtOH
Gly
Glu
Glu
PhenytoinGABA
ACh
NCAM1
GDNF
GFRA1
RET
FN
ITGBn
NGF
NTRK3
IGF1
IGFR
HGF
MET
PDGFB
PDGFR
EGF
ERBBn
KLF7
DOK5
CPNEn
CD44
MMP9
PZP
MMP10
Ptd Cho
IP2
PITP
SGM S1
PLC
DAG
IP3
PLCBn
ITPKB
INPPnx
IP4
NTN4
DCC
CORT
MRGPRX2
Gq11
GRK5
CNTNAP2
NRXN1
NXPH1
NRXN1
LAMxn
VOCC
CACNA1x
NMDA
GRIN2x
NLGNnx
Kainate
GRIK5
AMPA
GRIAn
DAG1
IRK
KCNJn
MaxiK
KCNMA1
VOPC
KCNE2
ATPase
ATP1B4
VOSC
SCNnx
GABAA
GABRB3
VSNL1
nAChR
α4β2
DAOA
DLG
DLG AP2
GKAP
SHANK
GRIP1
UBE3x
DMD
NCS1
FXYD6
K+
K+
Na+
Cl−
PTPRT
PTPRB
PTPRG
SHC
GRAP2
GRB2
GRB14
SOS
CNTN6
Ras
CALMn
Ca2+
RAPGEF2
ARHGAP48
RELN
VLDLR
ApoER2
DAB1
SEPT9
ARHGEF18
SORCS2
pro NGF
NGFR
SMAP2
GPR 56
Ga12/13
PLXN
SEMA
NRP
PI3K
TIAM
Raf
PEBP4
Rac
RASSFn
PARD3
MEK
PPP3Cx
OPHN1
ARHGAP32
EEF1A1
TRAPPC9
PP1
PP2A
RhoA
Cdc42
ERK
PPP1R14C
ER
ATXNn
IP3R
ITPRn
RYR
RYR3
PSEN
PSEN2
SERC
A2B
RBFOX1
FAT4
UGGT1
Glu
mt
SLC
25A12
Ca2+
Ca2+
Na+
STIM1
SOCC
ORAIn
PMCA
ATP2B2
NCX
SLC8An
Leak
NALCN
RBPJ
CREB
NCOR2
SEC16B
CREBBP
BAZ1A
NFATC2
EGR2
NFAT


AGP I
5.81
11
AGP II
5.58
2
4.81
3
4.44


AGP 2
AGP 1
5.07
4.81
3.33
2.76
4.89
1.88
−log10(p)
Recombination rate (cM/Mb)
0
2
4
6
20


AGP 2
AGP 1
3.50
5.56
5.64
6
4
2
0
−log10(p)
20
0
Recombination rate (cM/Mb)

Implications

Both phenotyping of AEs and GWAS have suffered from the use of less appropriate (univariate, linear model) statistical approaches.

As a novel computational biostat's approach, µGWAS/µPhene can integrate information from

- complex phenotypes and
- wide loci (several neighboring SNPs)

from a few hundred subjects only.

Data collected during Phase III can now be used to

- Identify genetic risk factors for AEs and Tx failure,
- Suggest novel directions for drug development, and
- Guide with selection of study populations,

References

Arbuthnot J (1710) An argument for divine providence taken from the constant regularity observ'd in the births of both sexes. *Philos T R Soc London* **27**:186-90

Hoeffding W (1948) A class of statistics with asymptotically normal distribution. *Ann Math Stat* **19**:293-325

Gehan EA (1965) A generalised two-sample Wilcoxon test for doubly censored samples. *Biometrika* **52**:650-3

Wei LJ, Knuiman MW (1987) A One-Sided Rank Test for Multivariate Censored Data. *Aust J Statistics* **29**:214-9

Wittkowski KM (1988) Friedman-type statistics and consistent multiple comparisons for unbalanced designs. *J Am Stat Assoc* **83**:1163-70, **87**:258

Wittkowski KM, Lee E et al. (2004) Combining several ordinal measures in clinical studies. *Stat Med* **23**:1579-92 (http://www.ncbi.nlm.nih.gov/pubmed/15122738)

Morales JF, Song T et al. (2008) Phenotyping genetic diseases using an extension of μ-scores for multivariate data. Stat *Appl Genet Mol* **7**:19 (http://hdl.handle.net/10209/490)

Papp KA, Fonjallaz P et al. (2008) Analytical approaches to reporting long-term clinical trial data. *Curr Med Res Opin* **24**:2001-8 (http://www.ncbi.nlm.nih.gov/pubmed/18534049)

Wittkowski KM, Song T et al. JE (2008) U-Scores for Multivariate Data in Sports. *J Quant Anal Sports* **4**:7 (http://hdl.handle.net/10209/492)

Diana M, Song T et al. (2009) Studying travel-related individual assessments and desires by combining hierarchically structured ordinal variables. *Transportation* **36**:187-206

Wittkowski KM, Song T (2010) Nonparametric methods for molecular biology. *Methods Mol Biol* **620**:105-53 (http://www.ncbi.nlm.nih.gov/pubmed/20652502)

Wittkowski KM, Sonakya V et al. (2013) From single-SNP to wide-locus: genome-wide association studies identifying functionally related genes and intragenic regions in small sample studies. *Pharmacogenomics* **14**:391-401 (http://dx.doi.org/10.2217/pgs.13.28)

The End

muStat.rockefeller.edu

µGWAS@Home