

# Inferring Causal Relationships Among Intermediate Phenotypes and Biomarkers: A Case Study of Rheumatoid Arthritis

*Wentian Li<sup>1</sup>, Mingyi Wang<sup>2</sup>, Patricia Irigoyen<sup>1</sup>, and Peter K. Gregersen<sup>1</sup>*

1. The Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research

2. College of Computer Science, Zhejiang University, Hangzhou, China

email: [wli@nslj-genetics.org](mailto:wli@nslj-genetics.org)

published in *Bioinformatics*, 22(12):1503-1507 (2006).

# One-Page Summary

The genetic data in Irigoyen et al *Arthritis & Rheumatism*, 52(12):3813-3818 (2005) presented at last year's retreat was reanalyzed using a new theoretical framework: **causal inference**.

Basically, causal inference is an attempt to assign **causal arrows** between correlated variables (e.g. rain  $\rightarrow$  dirt road, instead of dirt road  $\rightarrow$  rain).

Causal inference on observational data is not always possible. For any chance to assign a causal arrow between two variables, one must have information on **the third variable** . Thus we have the “causality triangle”.

The three variables we are analyzing are: **anti-cyclic citrullinated peptide (anti-CCP)**, **rheumatoid factor (RF)**, and **shared epitope (SE)**. The first two are auto-antibodies used as biomarkers for rheumatoid arthritis and the last one is risk alleles at HLA-DRB1 in MHC region for rheumatoid arthritis.

## Local Causality Rule

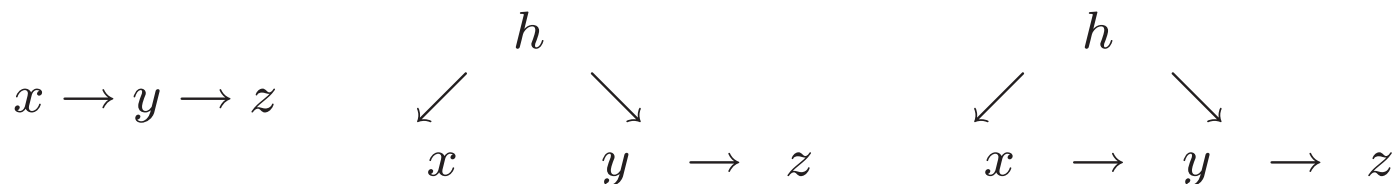
In an ideal situation, all relevant variables are collected and measured and any causal model can be tested by (e.g.) a likelihood ratio test. When we only have part of the data (“local”), it is impossible to calculate the likelihood.

A solution is to **exclude** causal models that are inconsistent with data. This “**local causality rule**” was discussed in Cooper *Data Mining and Knowledge Discovery*, 1:203-224 (1997).

Exclusion of inconsistent causal model may not lead to a single **unique** consistent model. However, in certain special cases, a unique causal relationship can be inferred.

# CCC Rule

The causality triangle consists of  $x, y, z$ . We assume that  $x$  is not caused by  $y$  and/or  $z$ , and a fourth hidden variable  $h$  is allowed as a potential confounding factor. Suppose  $x$  and  $y$  are correlated,  $y$  and  $z$  are correlated, **but  $x$  and  $z$  are uncorrelated conditional on  $y$** , then only the following causal relationships are possible:



In all three models, a unique causal arrow is assigned:  $y \rightarrow z$ .

CCC means correlated, correlated, correlated

# Our data on anti-CCP, RF, and SE

anti-CCP	RF	SE	No. samples
+	+	+	960
+	+	-	128
+	-	+	84
+	-	-	19
-	+	+	95
-	+	-	74
-	-	+	214
-	-	-	149

All are rheumatoid arthritis patients (N=1700), collected from NARAC (North American Rheumatoid Arthritis Consortium).

## Correlation and Conditional Correlation Analysis

	unconditional			
	$X^2$	OR 95%CI		$p$ -value
antiCCP,RF	661.7	22.7 (17.3, 29.8)		0
antiCCP,SE	190.7	5.1 (4.0, 6.5)		0
RF, SE	80.5	2.9 (2.3, 3.7)		$10^{-19}$
	conditional			
	$X_{s1}^2 + X_{s2}^2 = X^2$	OR <sub>1</sub> 95%CI	OR <sub>2</sub> 95%CI	$p$ -value
antiCCP,RF	3.9+0.4=4.3	1.7 (1, 2.9)	0.9 (0.6, 1.6)	0.12
antiCCP,SE	520.2+103.8=624.0	25.7 (18.5, 35.8)	13.6 (7.8, 23.7)	0
RF, SE	111.2+17.8=129.0	5.8 (4.1, 8.3)	3.1 (1.8, 5.3)	0

The most important observation is that **anti-CCP and RF are uncorrelated conditional on SE**

# CONCLUSION

anti-CCP  $\rightarrow$  RF

by replacing  $x$  with SE,  $y$  with anti-CCP, and  $z$  with RF

# DISCUSSION

- + Be aware of sample **selection bias**
- + **Affection status** is not one of the three variables (all samples are rheumatoid arthritis patients), so we are not inferring a cause of the disease
- + Anti-CCP autoantibody is **highly specific** for rheumatoid arthritis, whereas RF can be positive for other autoimmune diseases.
- + The causal inference procedure used here is for observational data with no time course. However, if **time series** is available, cause always precedes effect.