



Distance Weighted Discrimination

**Xingliang ‘Everett’ Zhou, J. Steve Marron
Saianand Balu, Xuxin Liu
LCCC, The University of North Carolina**

**Michael Nebozhyn, Louise Showe
The Wistar Institute**



NCI—caBIG –ICR –UNC--DWD

Why do we need DWD?

- **Small data sets are available at different institutes**
- **DWD can combine microarray data sets**
- **Performs systematic bias adjustment**
- **Improves statistical power by increasing sample size**
- **Outperforms conventional approaches**



NCI—caBIG –ICR –UNC--DWD

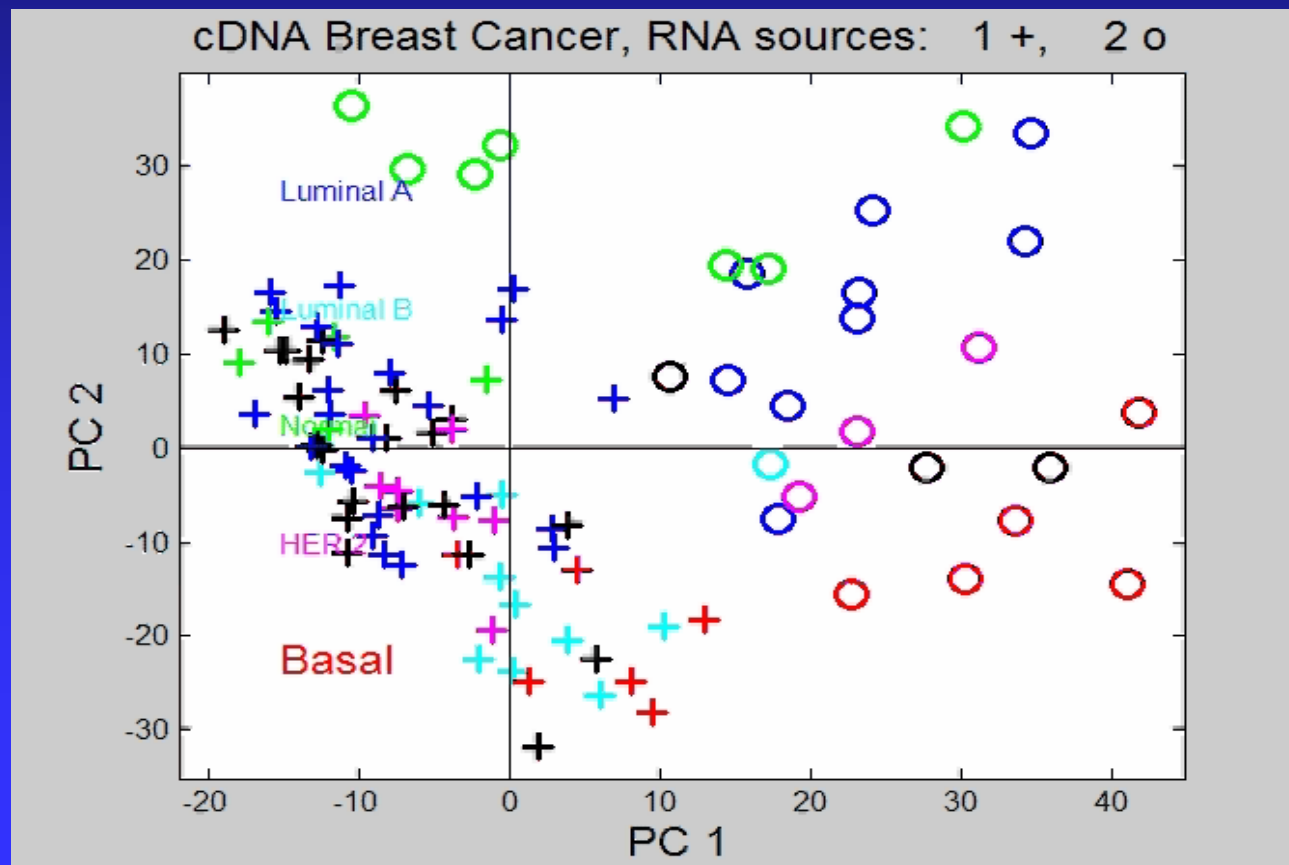
Current status of DWD

- **Functions as an analytical tool**
- **Input: Stanford format text delimited files and MAGE-ML format files**
- **Basic visual diagnostics**



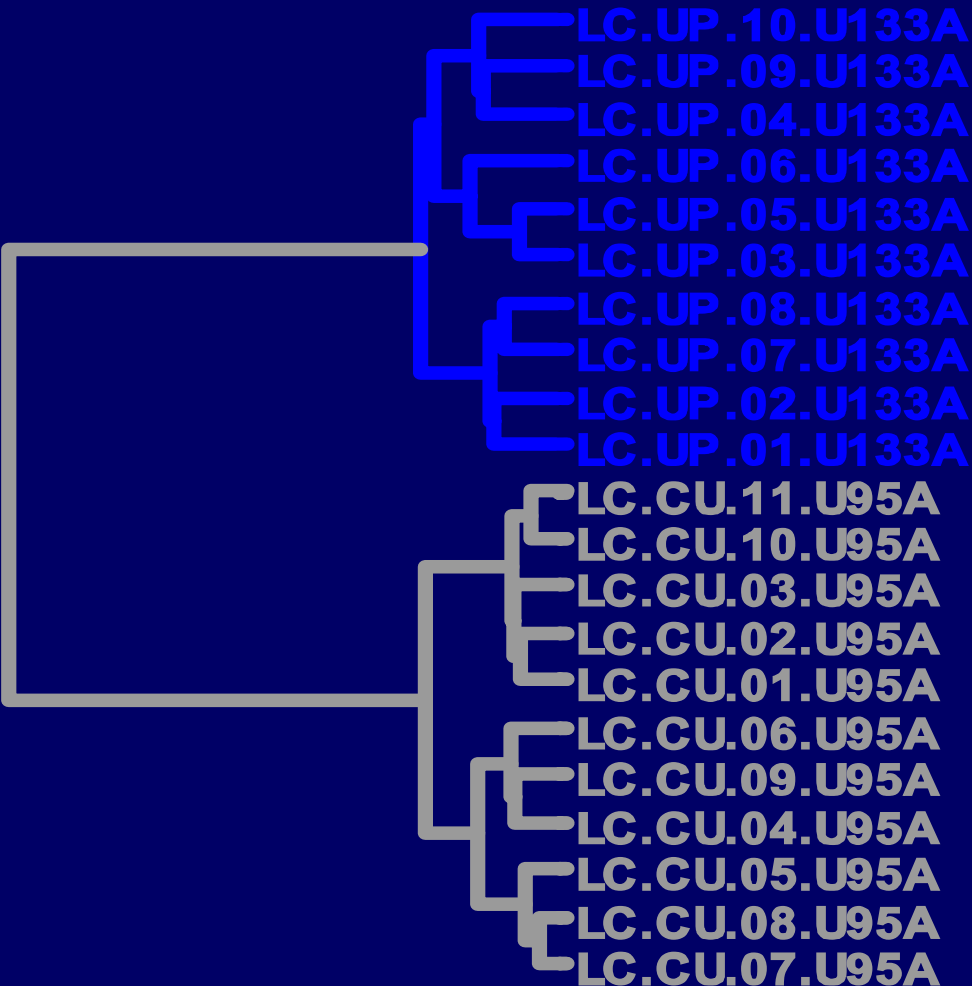
DWD Background

Purpose: Merge disparate Microarray Data Sets (e.g. Affy – cDNA)

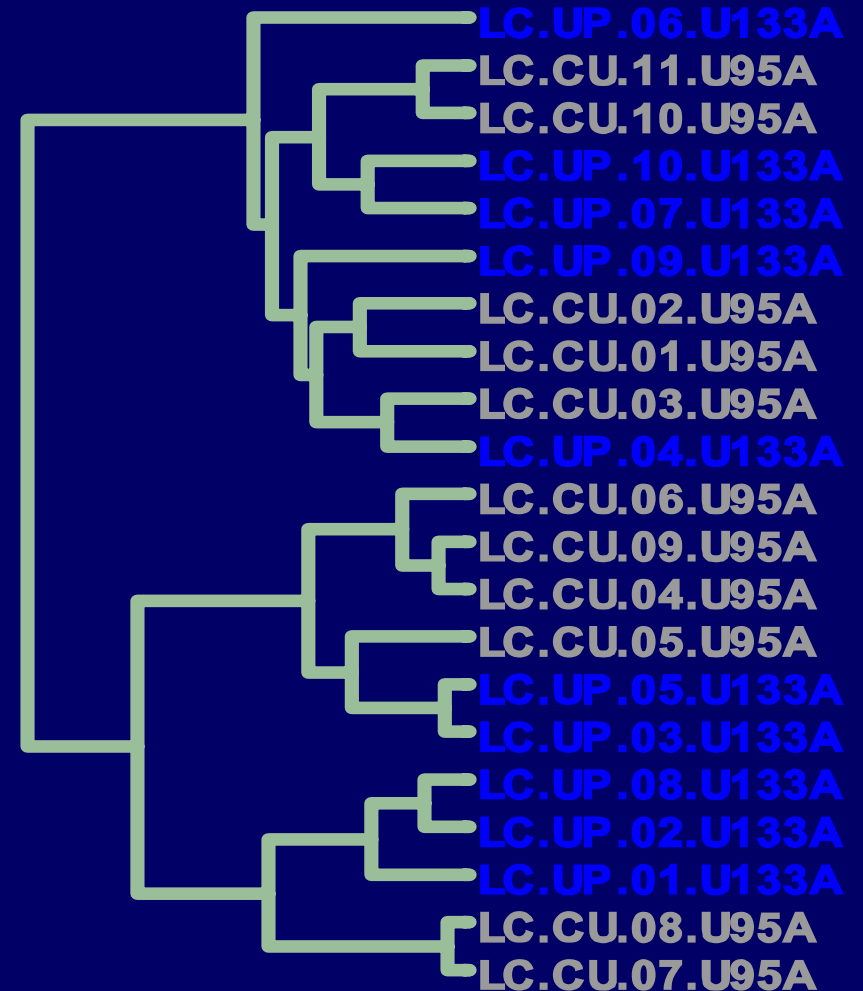


Merging LC.CU to LC.UP

Before



After



What Problems Did We Run into When Programming with MAGE-ML??

- **There are two different ways to get QuantitationTypes**
 - **QuantitationType_package**
Sum of QuantitationTypes
 - **BioAssayData**
QuantitationTypes associated with a particular BioAssayData
- **Output on a single line**
- **How to reassemble the components of BioAssayData into a BioAssayData?**

What Problems Did We Run into When Programming with MAGE-ML

Cont.

- What is the default name order if `bioDataCube.getNameOrder() == null` ?
DQB??
- DTD validation worked for Affymetrix, but failed for Agilent
- Lots of objects in MAGE-ML not stored in caDSR as CDEs

Future Plans for DWD

- **Integrate into caGRID as analytical service**
- **Improve the speed for visualization**
- **caBIG integrated training**
- **New features**
 - **Back classification (diagnostic)**
 - **Adjustment quality index (numeric summary performance)**



The End

