LUST at a Glance

- Input consists of **continuous** mRNA expression data for a set of patients whose tumors were biopsied, along with survival times from diagnosis.
- The expression data is binned into “overexpression,” “underexpression,” and “medium expression.” A new array of **discrete** data is made with enries of -1, 1 and 0, respectively.
- The **discrete** data is used to group genes that tend to overexpress and/or underexpress together. These groups are called *metagenes*.
- The **discrete** expression data is analyzed again to find smaller sets of genes that regulate metagene expression, these smaller sets are called *signatures*.
- A model θ is produced that takes **continuous** expression data for signatures as input and outputs a real number. Values of θ can be used to place patients in high and low risk categories.