



Instability

Research
Interest

Method

Results and
Conclusions

Systematic Normalization with Multiple Housekeeping Genes for the Discovery of Genetic Dependencies in Cancer

Oliver BONHAM-CARTER and Yee Mon THU
Allegheny College,
Meadville, PA

7th Sept 2019

Cancer from Instability

Genomic instability characterizes many cancers

Instability

Research Interest

Method

Results and Conclusions

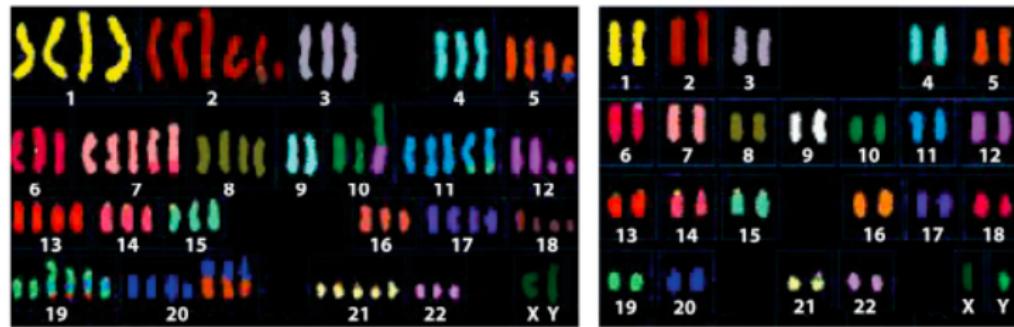


Figure: ?? Alberts, B. "Molecular biology of the cell Sixth edition Ch. 14, 755756." Garland Science, Taylor and Francis Group (2015).

- The karyotype of reduced stability of typical cancer cells, Figure left, and one of cancer cells with a relatively stable genome, Figure right.

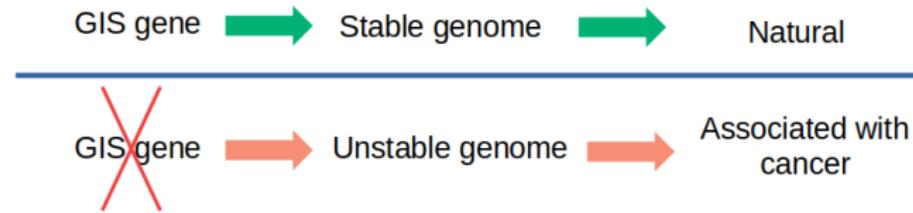
Instability From Genes

Instability

Research Interest

Method

Results and Conclusions



- GIS genes: Genes that suppress genome instability
- When GIS gene(s) is /are disabled, then an elevated risk of cancer association may arise in genome
- Instability seems to enable some cancers to survive

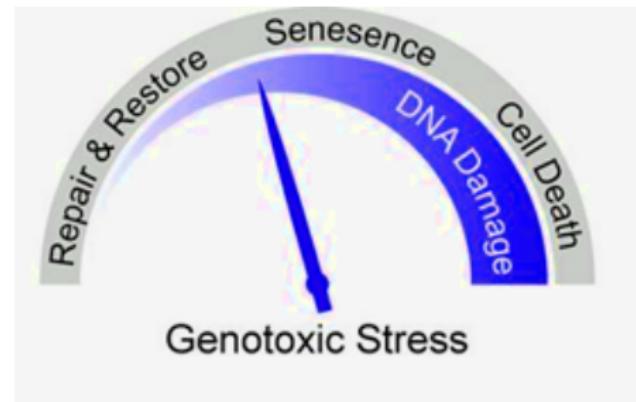
Equilibrium and Instability

Instability

Research Interest

Method

Results and Conclusions



- Relatively low levels of instability in the genome:
 - Repair and restore mechanisms function
 - Natural mutations to inspire genetic diversity
- High levels of instability:
 - Dysfunction, breakdown and *likely* cell death
 - Has cancer become dependent on some part of its unstable environment?

Correlations in Genome

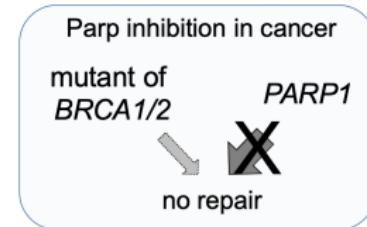
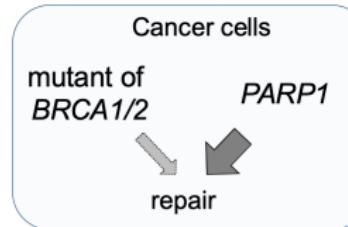
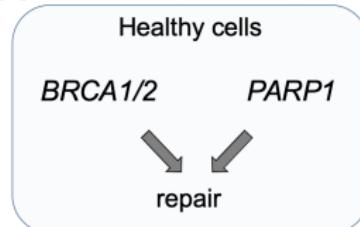
Instability

Research Interest

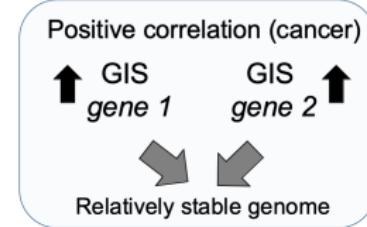
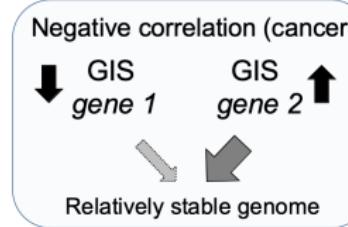
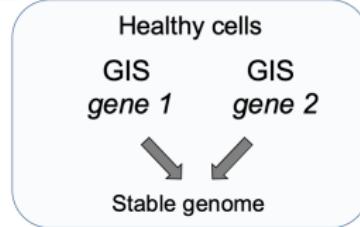
Method

Results and Conclusions

A



B



- **A.** Gene interactions introduce stability (*instability* in the genome)
- **B.** GIS genes may work together to gain stability.

Correlation Between Genes and Steps

Instability

Research
Interest

Method

Results and
Conclusions



- A large-scale approach to discovering correlations between the genes of our study groups
- R^2 values from linear regression for correlation
- Heatmaps to visualize the correlations for tests

Data and Selection

Instability

Research
Interest

Method

Normalization
and Models

Results and
Conclusions

- Genomic Data Commons Data Portal (National Cancer Institute).
- **Proof of concept:** Random selection of ten data sets of breast cancer gene expression
- Selected subset of genes limited search space, were specific to breast cancer research and helped to reduce noise in results
- Computational approach allows for scalability approach to determine correlations across large gene sets

Normalization

All values had to be normalized to enable comparisons

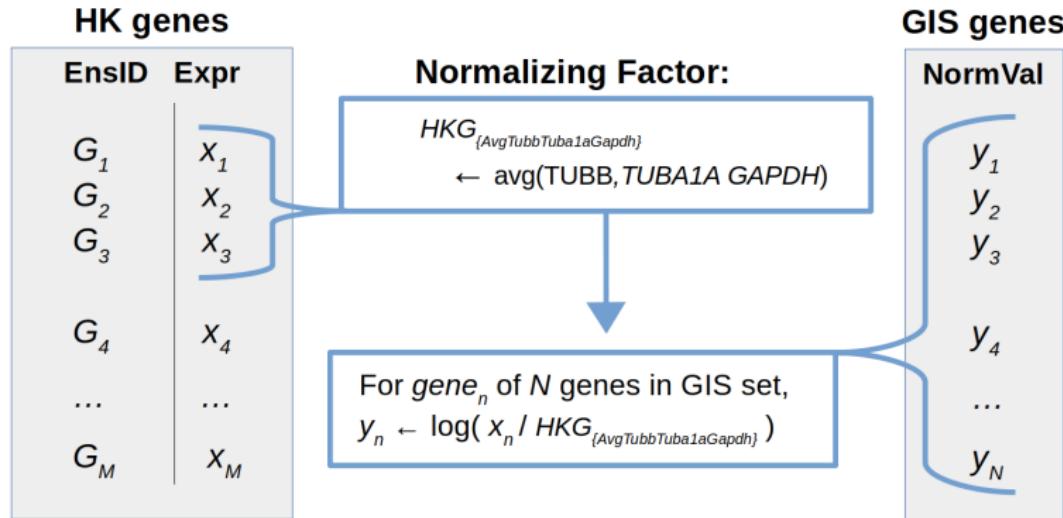
Instability

Research
Interest

Method

Normalization
and Models

Results and
Conclusions



- The expression values of the GIS genes were normalized by the averaged values of selected housekeeping gene expressions of same set.
- Automation done in R

Linear Models and R² Values

R code and flow

Instability

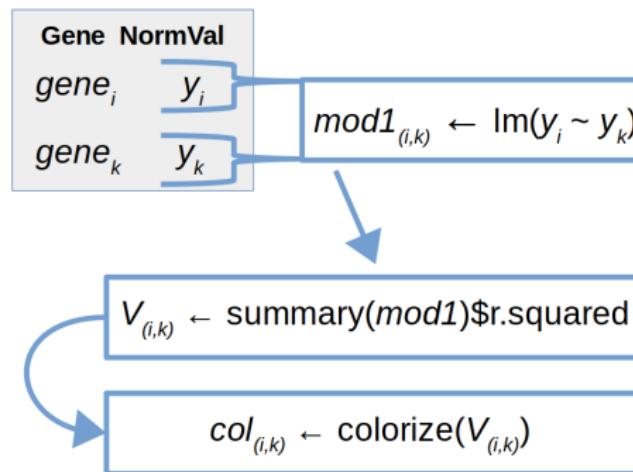
Research Interest

Method

Normalization and Models

Results and Conclusions

GIS genes



Heatmap

	$gene_k$	$gene_i$
$gene_k$	$col_{(i,k)}$	$col_{(i,i)}$
$gene_i$	$col_{(k,k)}$	$col_{(k,i)}$

- A set of normalized gene expressions was extracted for a particular gene from ten datasets of our trial. Each set of genes was regressed over all others in an *all-against-all* test
- To locate correlation between genes, we applied linear regression models and collected the R² values to place in heatmaps

Linear Models and R^2 Values

R code and flow

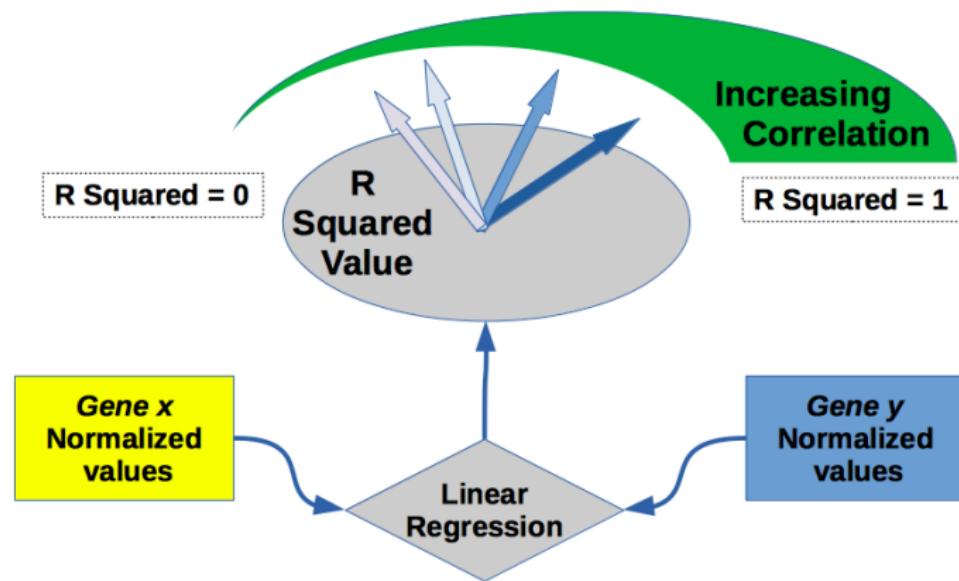
Instability

Research
Interest

Method

Normalization
and Models

Results and
Conclusions



- Scale [0,1]: High values of R^2 indicated strong correlations between values from each set of corresponding genes

Normalizing Techniques

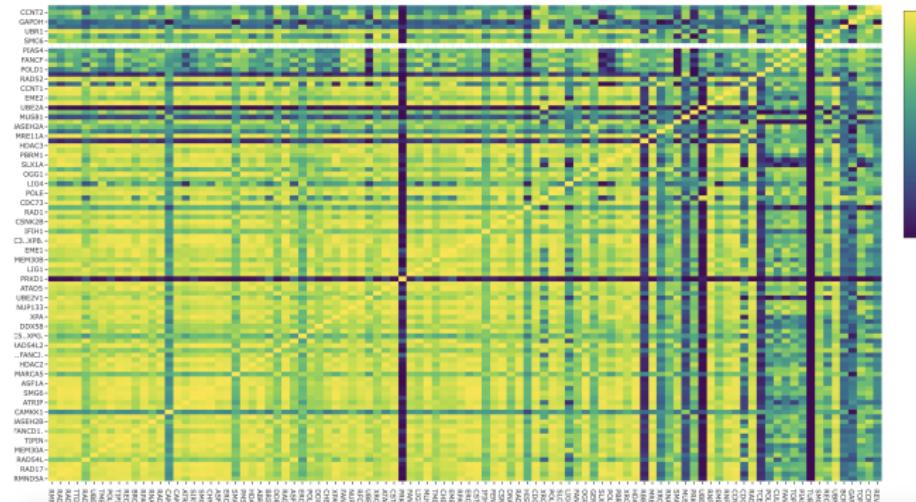
Few HK genes to normalize GIS genes

Instability

Research Interest

Method

Results and Conclusions



- The R^2 values from normalization where too few Housekeeping genes were used to normalize provided poor results.
- Biologically unlikely: Too many R^2 values to suggest high correlations

Normalizing Techniques

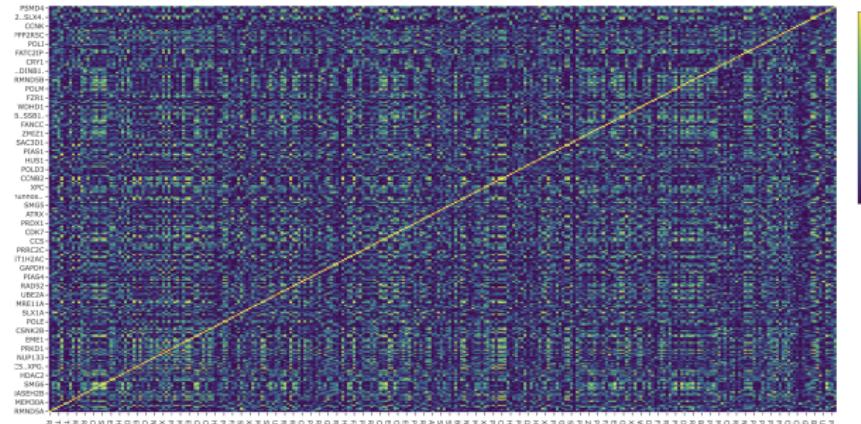
Ten HK genes to normalize GIS genes: Trial 1

Instability

Research
Interest

Method

Results and
Conclusions



- The R^2 values from normalization where ten housekeeping genes were used to normalize provided poor results.
- Biologically plausible; comparable with other sets of gene expressions, normalized by same ten housekeeping genes taken from their set.

Normalizing Techniques

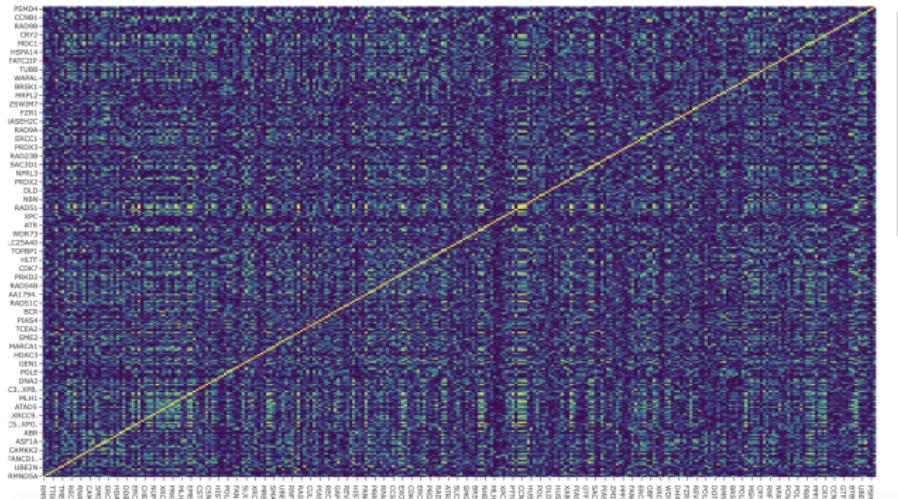
Ten HK genes to normalize GIS genes: Trial 2

Instability

Research
Interest

Method

Results and
Conclusions





Normalizing Techniques

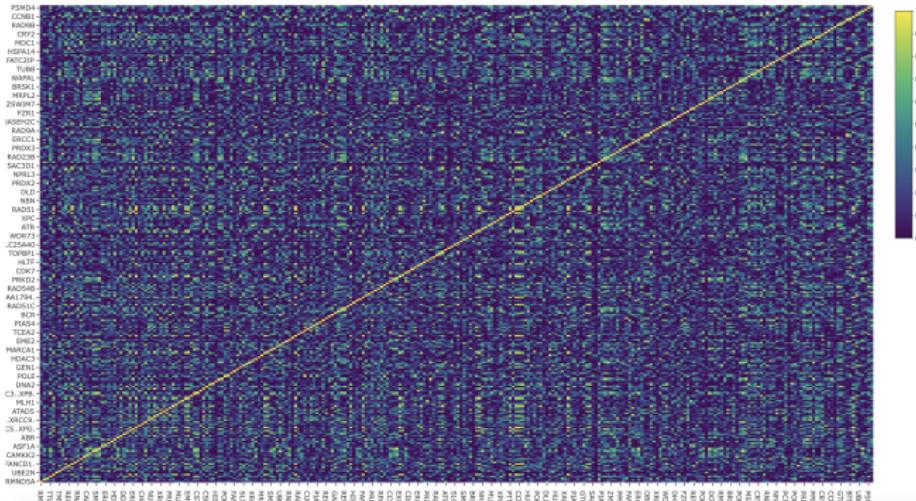
Ten HK genes to normalize GIS genes: Trial 3

Instability

Research Interest

Method

Results and Conclusions



Normalizing Techniques

Ten HK genes to normalize GIS genes: Trial 2

Instability

Research
Interest

Method

Results and
Conclusions

- **Single Expression Normalization:** Too few housekeeping genes used to normalize was not effective: correlations were biologically improbable.
- **Multiple Expression Normalization:** Multiple housekeeping genes used to normalize was successful in providing means to compare genes from multiple sets and to find biologically relevant correlations
- Our method was a pilot study, provided a good means to normalize data, and identified the co-expression of gene pairs in breast cancer tissues



Instability

Research
Interest

Method

Results and
Conclusions

Thank You! Questions?

obonhamcarter@allegheny.edu
ythu@allegheny.edu

<https://www.cs.allegheny.edu/sites/obonhamcarter/>