

GPU Data Mining in Neuroimaging Genomics

Bob Zigon

Beckman Coulter
Indianapolis, Indiana

May 10, 2017

- Background
- ANOVA for Voxels and SNPs
- VEGAS for Voxels and Genes
- High Speed GPU Monte-Carlo Simulator

What is Neuroimaging Genomics?

- 1 Neuroimaging Genomics is the fusion of brain imaging and genotyping data.
- 2 Study the influence of genetic variation on brain structure and function.

MRI and Sequencing data



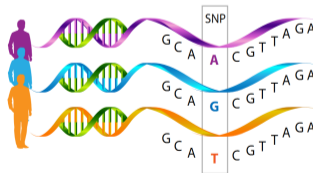
MRI instrument



MRI data



Genotyping instrument



Genotyping data

Develop an interactive tool for studying Alzheimer's Disease by coupling a 3D brain explorer with a genome explorer.

Prior Art

120 ROI's

20,000 SNP's

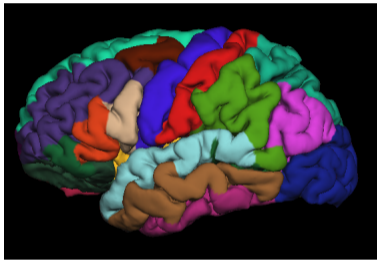
⇒

Our Goal

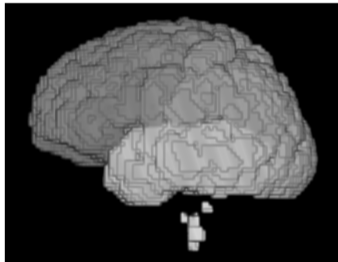
1,000,000 voxels

1,000,000 SNP's

Problem Definition



Brain with 120 Regions of Interest



Brain with 1,000,000 voxels

How We Do It – The UI

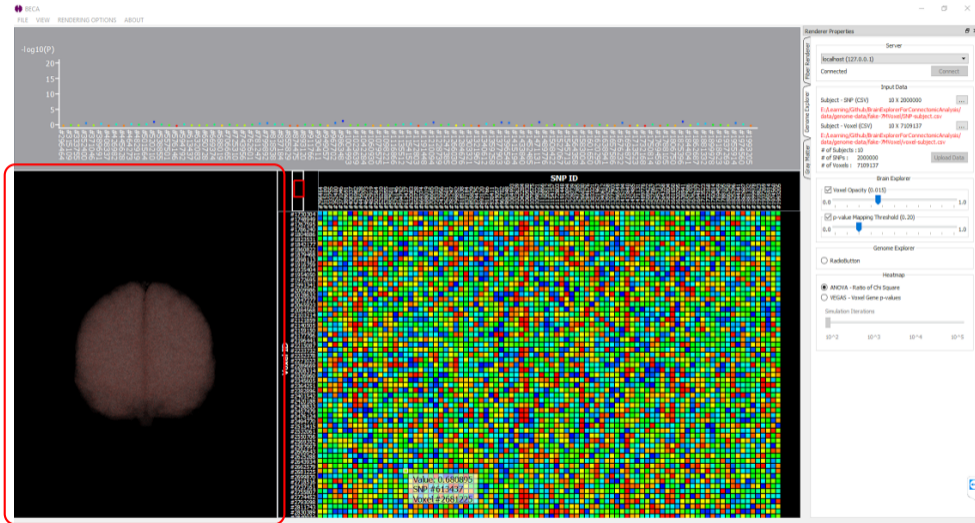
The screenshot displays the BECA software interface, which is used for brain connectivity analysis. The main window is divided into several panels:

- Top Panel:** A Manhattan plot showing $-\log_{10}(P)$ values for various SNPs. The y-axis ranges from 0 to 20. The x-axis lists SNP IDs such as rs1093205, rs1979560, rs1057348, and rs1919623.
- Left Panel:** A grayscale image of a brain slice, likely representing the anatomical structure being analyzed.
- Center Panel:** A large heatmap where the vertical axis is labeled "Voxel ID" and the horizontal axis is labeled "SNP ID". The heatmap cells are colored in a multi-color scale (red, yellow, green, blue, purple) representing different values. A small red box highlights a specific region in the top-left corner of the heatmap.
- Right Panel:** A "Render Properties" sidebar with several sections:
 - Server:** Shows the connection status to localhost (127.0.0.1) as "Connected".
 - Input Data:** Displays "Subject - SNP (CSV)" as 33 x 2000000 and "Subject - Voxel (CSV)" as 30 x 7109137. It includes links to GitHub repositories for data and analysis.
 - Brain Explorer:** Contains sliders for "Voxel Opacity (0.015)" and "p-value Mapping Threshold (0.20)".
 - Genome Explorer:** Includes a "Radefutton" option.
 - Heatmap:** Offers "ANCOVA - Ratio of Chi Square" and "VEGAS - Voxel Gene p-values" as visualization options, along with a "Simulation Iterations" slider.

At the bottom of the heatmap, a tooltip provides the following information:

- Value: 0.090805
- SNP: #6134137
- Voxel: #2681225

How We Do It – The UI



Brain Explorer

How We Do It – The UI

The screenshot displays the BECA (Brain Explorer and Connections Analyzer) software interface. At the top, the title bar reads "BECA" with menu options: FILE, VIEW, RENDERING OPTIONS, ABOUT. The main window is titled "SNP Explorer" and is divided into several panels:

- Top Panel:** A Manhattan plot showing $-\log_{10}(P)$ values for various SNPs. The y-axis ranges from 0 to 20. The x-axis lists SNP IDs such as rs1044392, rs1044393, etc.
- Left Panel:** A 3D brain model showing a coronal slice of the brain.
- Right Panel:** A heatmap where each cell's color represents a p-value for a specific SNP-voxel pair. The columns are labeled with SNP IDs and the rows with voxel IDs. A tooltip at the bottom right of the heatmap shows: "Value: 0.090805", "SNP: #6134437", "Voxel: #2681225".
- Far Right Panel:** "Render Properties" sidebar with settings for Server (localhost), Input Data (Subject - SNP (CSV) 33 x 2000000, Subject - Voxel (CSV) 30 x 7109137), Brain Explorer (Vessel Opacity: 0.015, p-value Mapping Threshold: 0.20), Genome Explorer (Radeflutton), and Heatmap (ANCOVA - Ratio of Chi Square, VEGAS - Voxel Gene p-values, Simulation Iterations).

SNP Explorer

Brain Explorer

How We Do It – The UI

BECA
FILE VIEW RENDERING OPTIONS ABOUT

SNP Explorer

$-\log_{10}(P)$

Brain Explorer

Heat Map

Render Properties

Server
localhost (127.0.0.1)
Connected

Input Data
Subject - SNP (CSV) 33 x 2000000
Subject - Voxel (CSV) 30 x 7109137

Brain Explorer
 Voxel Opacity (0.015)
 p-value Mapping Threshold (0.30)

Genome Explorer
 Radeflutton

Heatmap
 ANCOVA - Ratio of Chi Square
 VEGAS - Voxel Gene p-values

Simulation Iterations
10⁻² 10⁻³ 10⁻⁴ 10⁻⁵

Values: 0.090895
SNP: #6134137
Voxel: #2681225

Brain Explorer

Heat Map

ANOVA - Analysis of Variance

Understand the relationship between the gray matter density from the MRI and the SNP genotype.

"Did the combination happen by chance or not?"

ANOVA - Analysis of Variance

Understand the relationship between the gray matter density from the MRI and the SNP genotype.

"Did the combination happen by chance or not?"

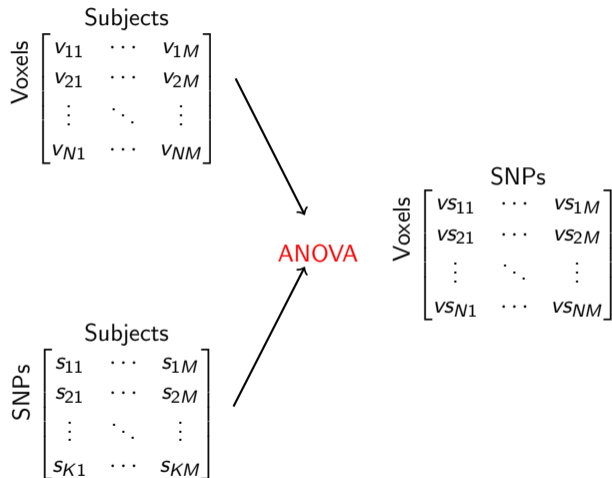
Computational complexity $\mathcal{O}(N_v * N_j * N_s)$

N_v - number of voxels

N_j - number of subjects

N_s - number of SNPS

ANOVA



VEGAS - VErsatile Gene based Association Study

Understand the relationship between the gray matter density from the MRI and the collective effect of multiple SNPs within a gene.

"Did the combination happen by chance or not?"

VEGAS - VErsatile Gene based Association Study

Understand the relationship between the gray matter density from the MRI and the collective effect of multiple SNPs within a gene.

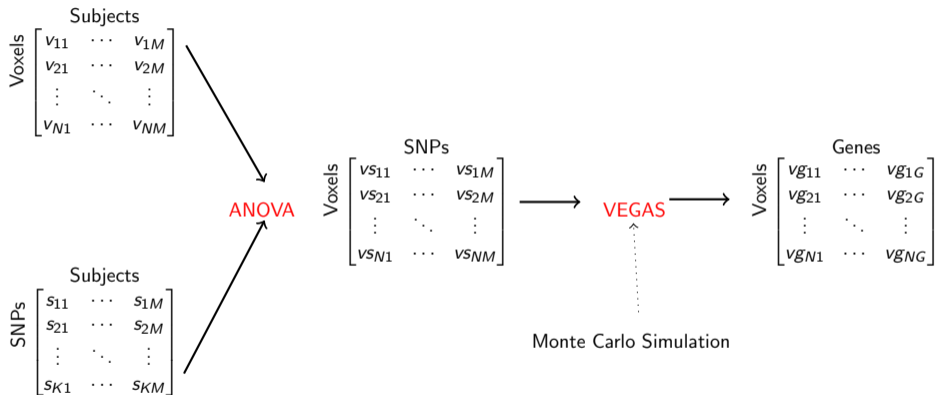
"Did the combination happen by chance or not?"

Computational complexity $\mathcal{O}(\underbrace{N_v * N_j * N_s}_{\text{ANOVA component}} + \underbrace{N_v * N_g * N_i}_{\text{Monte-Carlo component}})$

N_v - number of voxels

N_g - number of genes

N_i - number of Monte-Carlo iterations ($10^2, 10^3, 10^4, 10^5$, or 10^6)



Video of the Interactive Neuroimaging Genomic Browser

How do you build a high speed Monte-Carlo Simulator
for an N dimensional problem on a GPU?

One Dimensional

for $i = 1$ to K **do**

 Choose X from $N(0,1)$

$Y = F(X)$

 Make decision about Y

end

One Dimensional

```
for  $i = 1$  to  $K$  do  
  | Choose  $X$  from  $N(0,1)$   
  |  $Y = F(X)$   
  | Make decision about  $Y$   
end
```

N-Dimensional

```
for  $i = 1$  to  $K$  do  
  | Choose  $n$  values from  $N(0,1)$  giving  $X^n$   
  |  $Y^n = F(X^n)$   
  | Make decision about  $Y^n$   
end
```

First Attempt at N-Dimensional

```
foreach voxel  $V$  in parallel do
  |
  foreach gene  $G$  in parallel do
    |
    for  $i = 1$  to  $K$  do
      |
      Choose  $n$  values from  $N(0,1)$  giving  $X^n$ 
       $Y^n = F(X^n)$ 
      Make decision about  $Y^n$ 
    end
  end
end
end
```

First Attempt at N-Dimensional

```

foreach voxel  $V$  in parallel do
  |
  foreach gene  $G$  in parallel do
  | |
  | | for  $i = 1$  to  $K$  do
  | | | Choose  $n$  values from  $N(0,1)$  giving  $X^n$ 
  | | |  $Y^n = F(X^n)$ 
  | | | Make decision about  $Y^n$ 
  | | end
  | end
end

```

	Slow	Fast	Theoretical
Memory Bandwidth (gb/sec)	20		320
GFLOPS	20		10,000

N-Dimensional Ah-Ha

In parallel, generate $n \times K$ values from $N(0,1)$ giving $X^{n \times K}$

$$Y^{n \times K} = F^{n \times n} \times X^{n \times K}$$

In parallel, decide about $Y^{n \times K}$

Slow N-Dimensional

```
foreach voxel  $V$  in parallel do
  foreach gene  $G$  in parallel do
    for  $i = 1$  to  $K$  do
       $n$  values from  $N(0,1)$ 
       $Y^n = F(X^n)$ 
      Make decision about  $Y^n$ 
    end
  end
end
```

Fast N-Dimensional

```
foreach voxel  $V$  sequentially do
  foreach gene  $G$  sequentially do
    In parallel, generate  $nK$  values
     $Y^{n \times K} = F^{n \times n} \times X^{n \times K}$ 
    In parallel, decide about  $Y^{n \times K}$ 
  end
end
```


Slow N-Dimensional

```
foreach voxel  $V$  in parallel do
  foreach gene  $G$  in parallel do
    for  $i = 1$  to  $K$  do
       $n$  values from  $N(0,1)$ 
       $Y^n = F(X^n)$ 
      Make decision about  $Y^n$ 
    end
  end
end
end
```

Fast N-Dimensional

```
foreach voxel  $V$  sequentially do
  foreach gene  $G$  sequentially do
    In parallel, generate  $nK$  values
     $Y^{n \times K} = F^{n \times n} \times X^{n \times K}$ 
    In parallel, decide about  $Y^{n \times K}$ 
  end
end
```

Slow N-Dimensional

```

foreach voxel V in parallel do
  |
  foreach gene G in parallel do
  | |
  | | for  $i = 1$  to  $K$  do
  | | |
  | | |  $n$  values from  $N(0,1)$ 
  | | |  $Y^n = F(X^n)$ 
  | | | Make decision about  $Y^n$ 
  | | end
  | end
end
  
```

Fast N-Dimensional

```

foreach voxel V sequentially do
  |
  foreach gene G sequentially do
  | |
  | | In parallel, generate  $nK$  values
  | |  $Y^{n \times K} = F^{n \times n} \times X^{n \times K}$ 
  | | In parallel, decide about  $Y^{n \times K}$ 
  | end
end
  
```

	Slow	Fast	Theoretical
Memory Bandwidth (gb/sec)	20	155	320
GFLOPS	20	2,000	10,000

800X improvement!

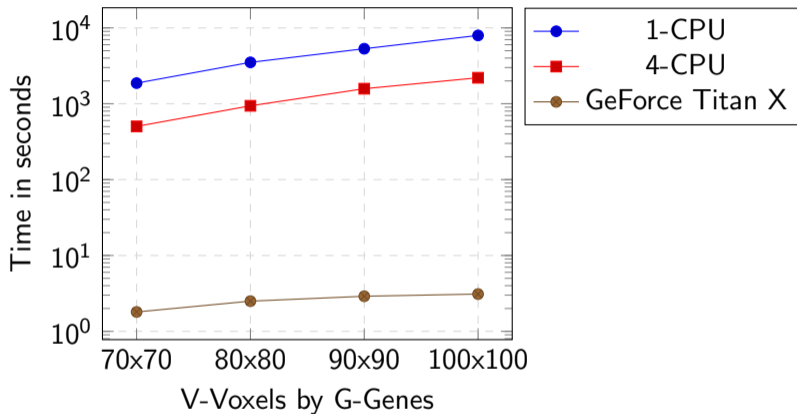


Figure 1: Execution times for 1 VEGAS run with $K=10,000$ Monte-Carlo iterations

Acknowledgements

- ① Professor Li Shen, Dept. of Radiology and Imaging Sciences, IU School of Medicine, NIH R01 EB022574, R01 LM011360, U01 AG024904, and IUPUI ITDP Program.
- ② Professor Shiaofen Fang, Computer Science Dept. Chair, Indiana University-Purdue University of Indianapolis
- ③ Professor Mohammad Al Hasan, Associate Professor Computer Science, Indiana University-Purdue University of Indianapolis
- ④ Huang Li, PhD Candidate, Computer Science, Indiana University-Purdue University of Indianapolis

Thank you

robert.zigon@beckman.com

