GPU Data Mining in Neuroimaging Genomics

Bob Zigon

Beckman Coulter
Indianapolis, Indiana

May 10, 2017

Outline

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

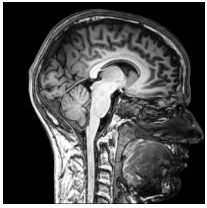
What is Neuroimaging Genomics?

1. Neuroimaging Genomics is the fusion of brain imaging and genotyping data.

2. Study the influence of genetic variation on brain structure and function.

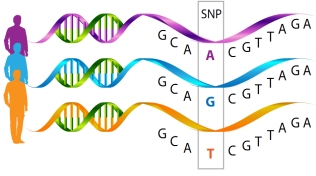
MRI and Sequencing data

MRI instrument



MRI data

Genotyping instrument



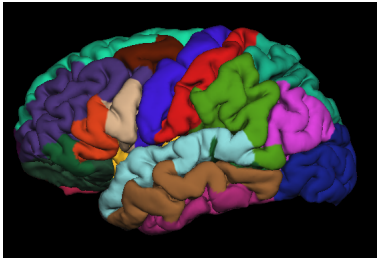
Genotyping data

Problem Definition

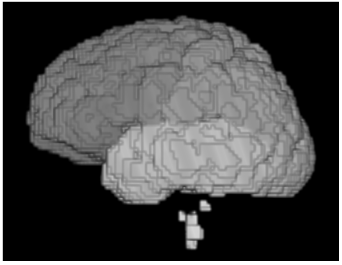
Develop an interactive tool for studying Alzheimer's Disease by coupling a 3D brain explorer with a genome explorer.

Prior Art		**Our Goal**
120 ROI's	$\Rightarrow$	1,000,000 voxels
20,000 SNP's		1,000,000 SNP's

Problem Definition

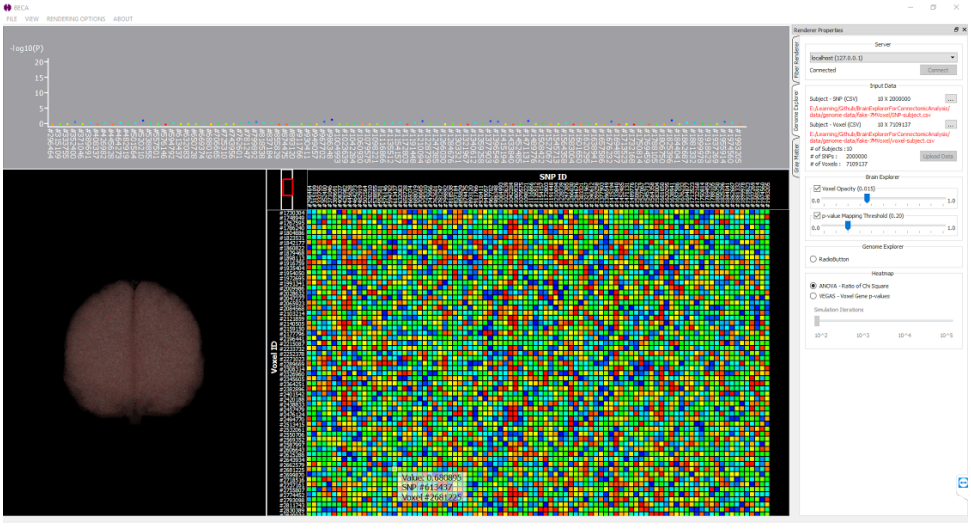


Brain with 120 Regions of Interest

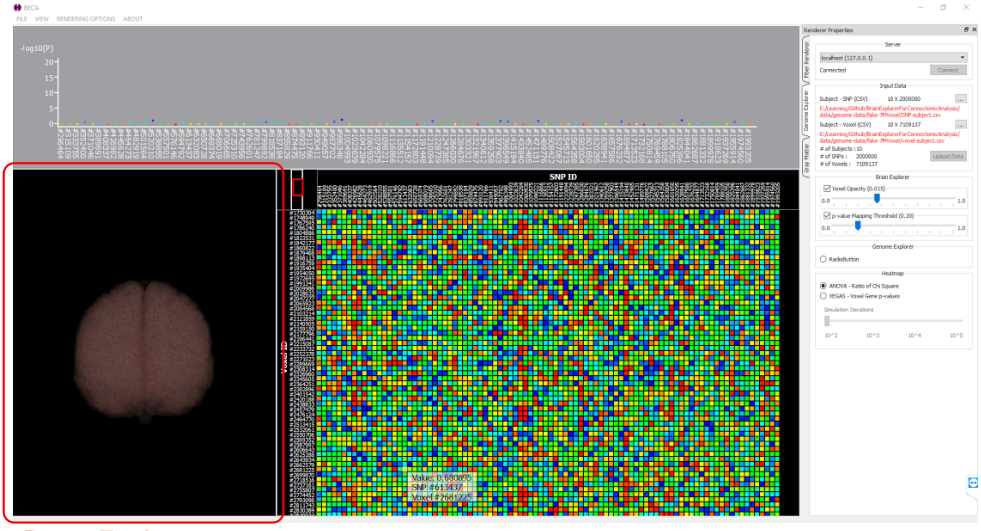


Brain with 1,000,000 voxels

How We Do It – The UI

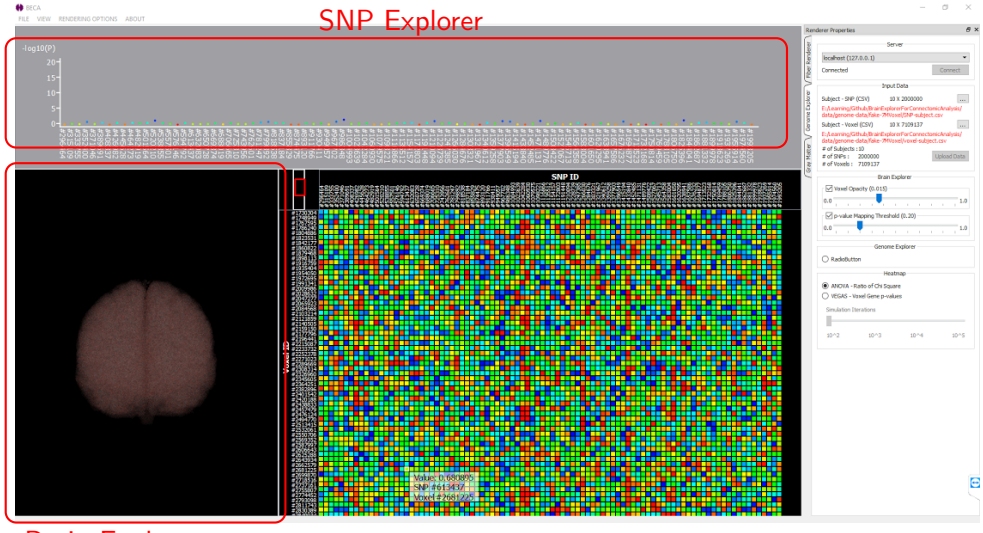


How We Do It – The UI

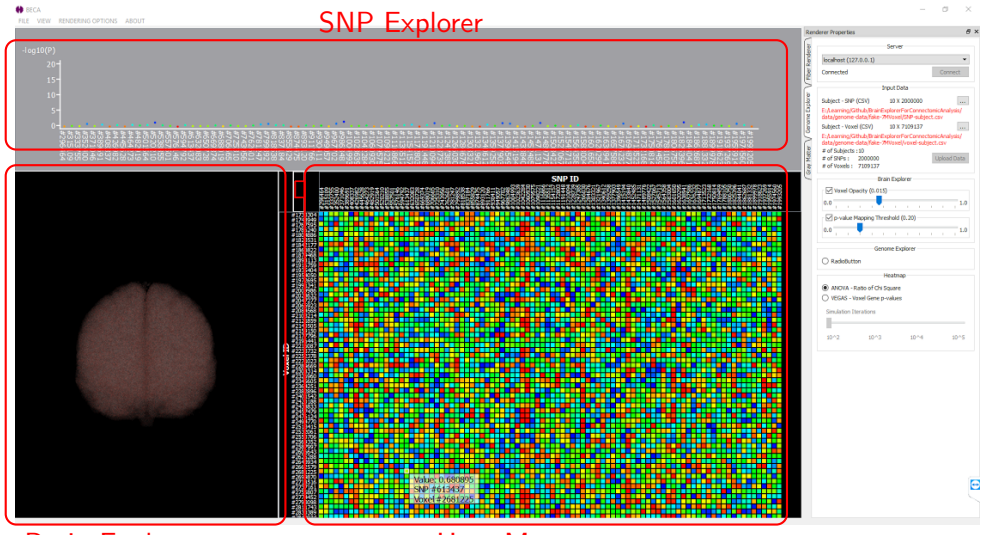


Brain Explorer

How We Do It – The UI



How We Do It – The UI



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

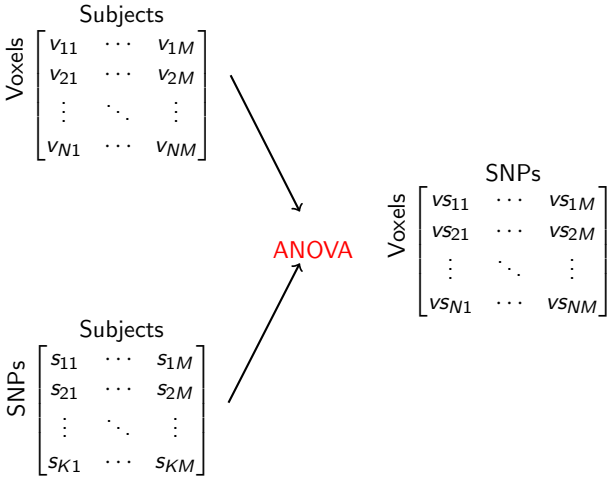
Understand the relationship between the gray matter density from the MRI and the SNP genotype.

"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


Subjects
Voxels
v_{11} $\cdots$ v_{1M}
v_{21} $\cdots$ v_{2M}
v_{N1} $\cdots$ v_{NM}
Subjects
SNPs
s_{11} $\cdots$ s_{1M}
s_{21} $\cdots$ s_{2M}
s_{K1} $\cdots$ s_{KM}
ANOVA
SNPs
Voxels
vs_{11} $\cdots$ vs_{1M}
vs_{21} $\cdots$ vs_{2M}
vs_{N1} $\cdots$ vs_{NM}

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

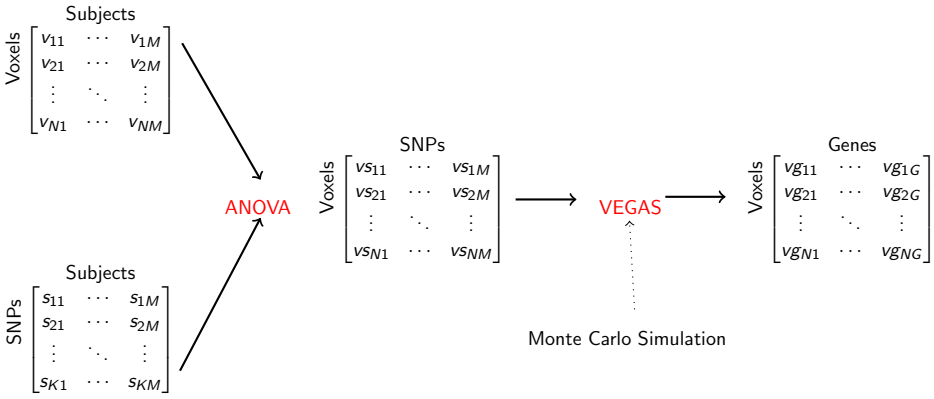
Understand the relationship between the gray matter density from the MRI and the collective effect of multiple SNPs within a gene.

"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)

VEGAS


Subjects
Voxels
$\begin{bmatrix} v_{11} & \cdots & v_{1M} \\ v_{21} & \cdots & v_{2M} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{NM} \end{bmatrix}$
Subjects
SNPs
$\begin{bmatrix} s_{11} & \cdots & s_{1M} \\ s_{21} & \cdots & s_{2M} \\ \vdots & \ddots & \vdots \\ s_{K1} & \cdots & s_{KM} \end{bmatrix}$
ANOVA
SNPs
Voxels
$\begin{bmatrix} vs_{11} & \cdots & vs_{1M} \\ vs_{21} & \cdots & vs_{2M} \\ \vdots & \ddots & \vdots \\ vs_{N1} & \cdots & vs_{NM} \end{bmatrix}$
VEGAS
Genes
Voxels
$\begin{bmatrix} vg_{11} & \cdots & vg_{1G} \\ vg_{21} & \cdots & vg_{2G} \\ \vdots & \ddots & \vdots \\ vg_{N1} & \cdots & vg_{NG} \end{bmatrix}$
Monte Carlo Simulation

Video

Video of the Interactive Neuroimaging Genomic Browser

Lessons Learned

How do you build a high speed Monte-Carlo Simulator

for an N dimensional problem on a GPU?

Lessons Learned

One Dimensional

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

Choose X from N(0,1)

$Y = F(X)$

Make decision about Y

end

Lessons Learned

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

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

Lessons Learned - Solution

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 $Y^{n \times K}$

Lessons Learned

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

Lessons Learned

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

Lessons Learned

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!

Vegas Results

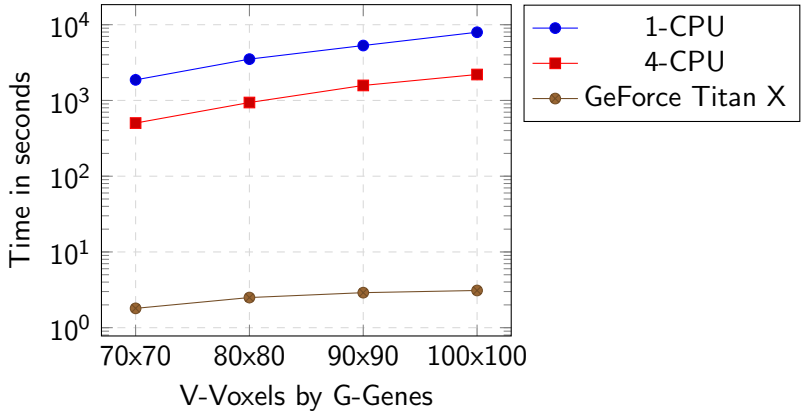


Figure 1: Execution times for 1 VEGAS run with K=10,000 Monte-Carlo iterations

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