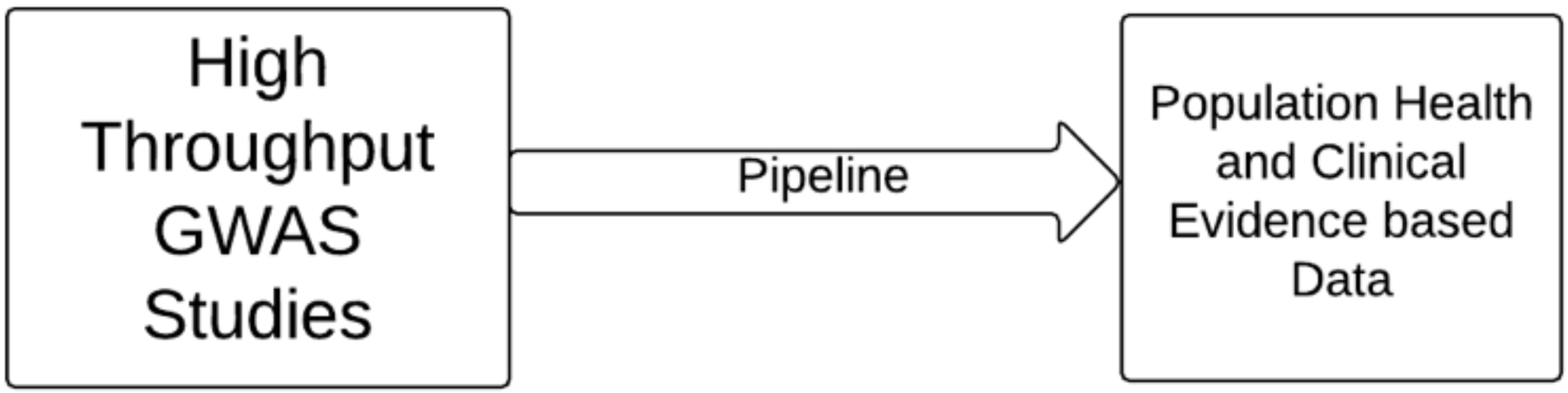
Work in Progress: Developing Polygenic disease scores from literature using ADNI

Arindam Basu
University of Canterbury
arindam.basu@canterbury.ac.nz
Presented at the QIMR Berghofer Research Meeting, Brisbane, 19th April, 2016

Challenge for Translational Research



Gene and SNP based data

Clinical and Aggregate level data

Challenge

- Construct a predictive model using SNP data from GWAS studies to predict and model diseases and health outcomes? [Genetic Prediction for Illness]

Proposed Steps

Identify a list of disease conditions/outcomes

↓

Prioritise or Weight the Diseases/Outcomes

↓

Identify Relevant SNPs from another source

↓

Develop Polygenic Risk Scores for the Diseases

Identification of diseases

- Based on the NHGRI/EBI GWAS study database
- Downloaded the entire data set
- Filtered the data base to have only Europeans, p-values 1e-8, disease conditions (not risk factors or measurements on risk or exposure variables)
- Selected groups of diseases (cardiovascular, immune system related, cancers, etc)

Entire GWAS catalog

GWAS Catalog | Home | Search | Diagram | **Download** | Help | About | EMBL-EBI | NIH National Human Genome Research Institute

Examples: breast cancer, rs7329174, Yang, 2q37.1, HBS1L

Home / Documentation / Downloads

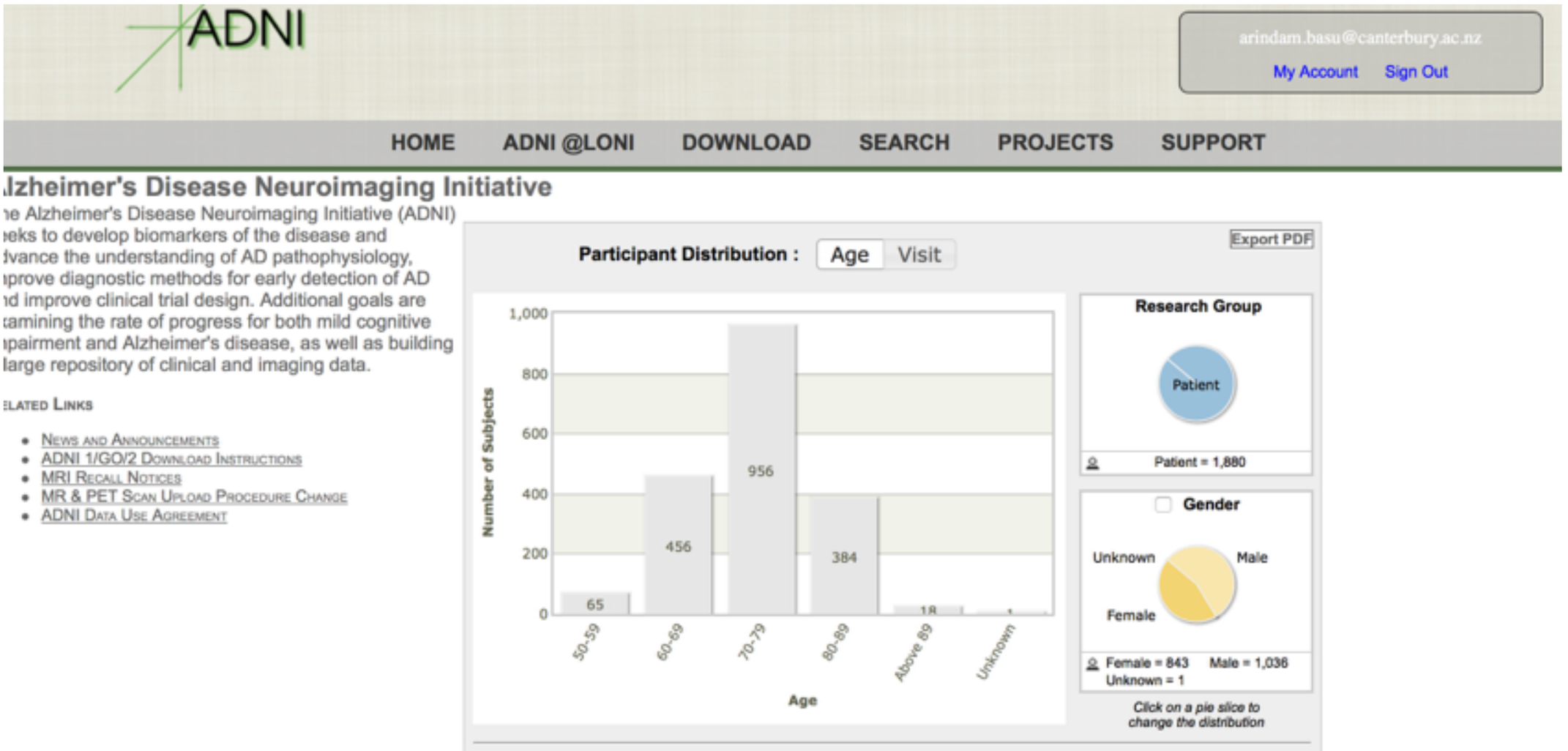
Downloading the GWAS Catalog

Description	Download Link	Format	Column header descriptions
The most recent edition of all the associations in the GWAS Catalog	Click to download	tab separated file	Click to view
The most recent edition of all the associations in the GWAS Catalog **with added ontology annotations**	Click to download	tab separated file	Click to view
The most recent edition of all the studies in the GWAS Catalog	Click to download	tab separated file	Click to view
The most recent edition of all the studies in the GWAS Catalog **with added ontology annotations**	Click to download	tab separated file	Click to view

Screenshot of CVD

dis_trait	chr_region	chromosome	snp1	allele	allele_freq	log_p_value	prevalence
Brugada syndrome	3p22.2	3	rs10428132	T	0.41	68	0.0002
Brugada syndrome	6q22.31	6	rs9388451	C	0.5	16.301	0.0002
Brugada syndrome	3p22.2	3	rs11708996	C	0.15	14	0.0002
Abdominal aortic aneurysm	9q33.2	9	rs7025486	A	0.23	9.301	0.01
Abdominal aortic aneurysm	12q13.3	12	rs1466535	C	0.58	9.301	0.01
Aortic-valve calcification	6q25.3	6	rs10455872	G	0.07	10.523	0.18
Atrial fibrillation	4q25	4	rs6817105	C	0.13	73.699	0.09
Atrial fibrillation	4q25	4	rs6843082	G	0.26	27.523	0.09
Atrial fibrillation	1q21.3	1	rs13376333	T	0.3	20.699	0.09
Atrial fibrillation	15q24.1	15	rs7164883	G	0.16	16.523	0.09
Atrial fibrillation	16q22.3	16	rs2106261	T	0.176	15.523	0.09
Atrial fibrillation	1q21.3	1	rs6666258	C	0.3	13.699	0.09
Atrial fibrillation	1q24.2	1	rs3903239	G	0.45	13.097	0.09
Atrial fibrillation	14q23.2	14	rs1152591	A	0.48	12.222	0.09

Alzheimer's Disease Neuroimaging Initiative

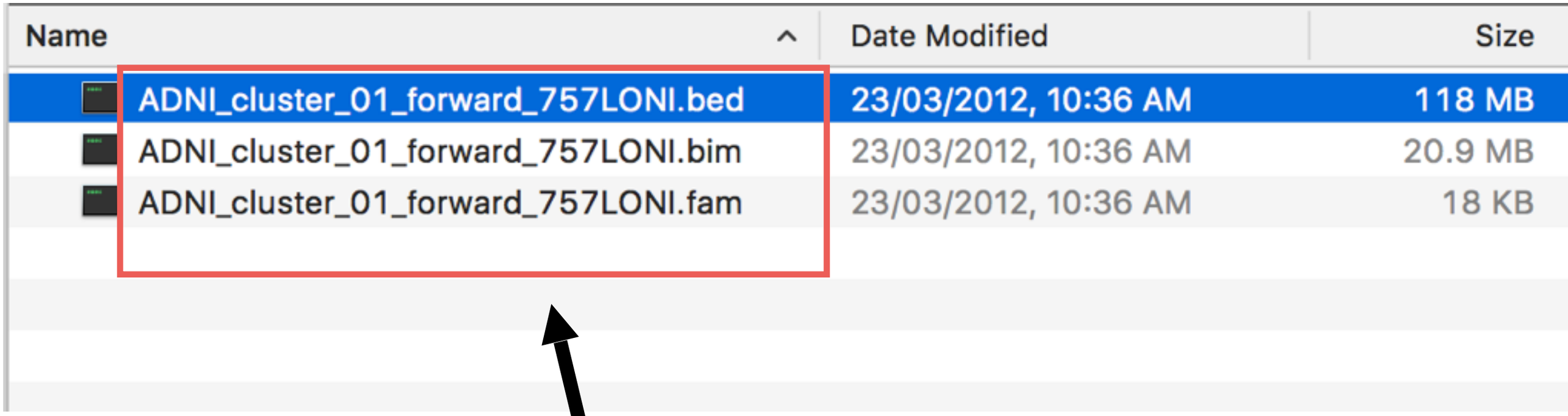


Sample Medical History File from ADNI data set

Table 1

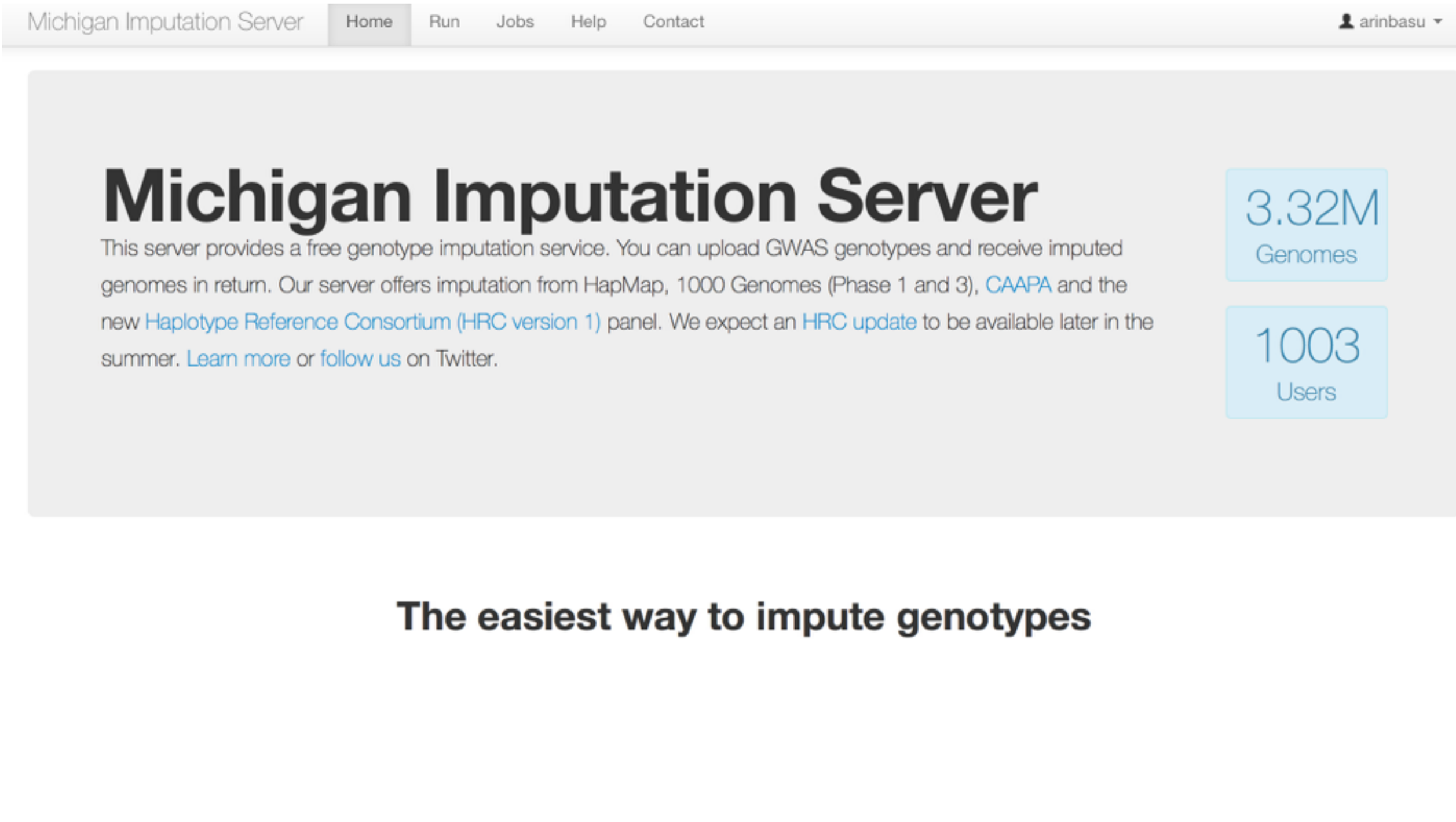
Phase	ID	RID	SITEID	VISCODE	VISCODE2	USERDATE	USERDATE2	RECNO	EXAMDATE	MHNUM	MHDESC
ADNI1	12	2	107	sc	sc	2005-08-17		1	2005-08-17	9	--/--/1984 Pancreatitis
ADNI1	14	2	107	sc	sc	2005-08-17		2	2005-08-17	9	--/--/2001 Non Insulin dependent diabetes, diet and exercise controlled.
ADNI1	16	2	107	sc	sc	2005-08-17		3	2005-08-17	3	--/--/1993 cataracts removed from both eyes
ADNI1	18	1	10	f	f	2005-08-18		1	2005-08-18	3	astigmatism OU since 1950, corrective lenses worn. Floaters in both eyes since 2003. Tonsille
ADNI1	20	1	10	f	f	2005-08-18		2	2005-08-18	6	Hepatitis A 1990
ADNI1	22	1	10	f	f	2005-08-18		3	2005-08-18	19	Lyme disease on 3 occasions: bull's eye lesion, otherwise asymptomatic, treated with oral AB's
ADNI1	24	3	107	sc	sc	2005-08-18		1	2005-08-18	4	1995-hypertension; 1995-hyperlipidemia
ADNI1	26	3	107	sc	sc	2005-08-18		2	2005-08-18	8	2000-osteoarthritis
ADNI1	28	3	107	sc	sc	2005-08-18		3	2005-08-18	9	1985-hyperthyroidism
ADNI1	30	3	107	sc	sc	2005-08-18		4	2005-08-18	12	1990-nephrectomy
ADNI1	32	3	107	sc	sc	2005-08-18		5	2005-08-18	16	1 ppd x 20 years, quit 1985
ADNI1	34	3	107	sc	sc	2005-08-18		6	2005-08-18	17	1995 prostate cancer; 1990 kidney cancer
ADNI1	36	3	107	sc	sc	2005-08-18		7	2005-08-18	18	1990 (L) total knee replacement, 1990 nephrectomy; 1995 prostatectomy
ADNI1	38	4	10	sc	sc	2005-08-18		1	2005-08-18	8	Osteoarthritis in both knees. Piroxicam 20mg daily. Problem started three years ago but stable.
ADNI1	40	4	10	sc	sc	2005-08-18		2	2005-08-18	12	History of benign prostatic hyperplasia since 1997. Doxazosin 2mg daily. Finding of increased
ADNI1	42	4	10	sc	sc	2005-08-18		3	2005-08-18	12	History of proteinuria of unknown cause since tem years ago.
ADNI1	44	4	10	sc	sc	2005-08-18		4	2005-08-18	16	Currently smokes ten cigarrets a day since 50 years ago.

ADNI Data in PLINK Format



The PLINK formatted data set from ADNI

Running imputation



Weighting?

dis_trait	chr_region	chromosome	snp1	allele	allele_freq	log_p_value	prevalence
Brugada syndrome	3p22.2	3	rs10428132	T	0.41	68	0.0002
Brugada syndrome	6q22.31	6	rs9388451	C	0.5	16.301	0.0002
Brugada syndrome	3p22.2	3	rs11708996	C	0.15	14	0.0002
Abdominal aortic aneurysm	9q33.2	9	rs7025486	A	0.23	9.301	0.01
Abdominal aortic aneurysm	12q13.3	12	rs1466535	C	0.58	9.301	0.01
Aortic-valve calcification	6q25.3	6	rs10455872	G	0.07	10.523	0.18
Atrial fibrillation	4q25	4	rs6817105	C	0.13	73.699	0.09
Atrial fibrillation	4q25	4	rs6843082	G	0.26	27.523	0.09
Atrial fibrillation	1q21.3	1	rs13376333	T	0.3	20.699	0.09
Atrial fibrillation	15q24.1	15	rs7164883	G	0.16	16.523	0.09
Atrial fibrillation	16q22.3	16	rs2106261	T	0.176	15.523	0.09
Atrial fibrillation	1q21.3	1	rs6666258	C	0.3	13.699	0.09
Atrial fibrillation	1q24.2	1	rs3903239	G	0.45	13.097	0.09
Atrial fibrillation	14q23.2	14	rs1152591	A	0.48	12.222	0.09

Next Steps

- Identify the SNPs
- Run regression with Cardiovascular disease in ADNI data set with selected SNPs and demographic and other risk variables
- Treat multiple disease conditions as a group
- Outcome variable?