

ADNI Data and Publications: A Unique Model of Open Data Access

Robert C. Green, MD, MPH

Associate Director for Research,

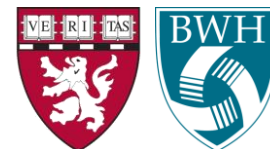
Partners Center for Personalized Genetic Medicine

Division of Genetics, Department of Medicine

Brigham and Women's Hospital, Harvard Medical School



CENTER FOR PERSONALIZED
GENETIC MEDICINE



NIH Grant Funding

U01 HG006500 (Green)

U01 AG024904 (Weiner)

R01 HG002213 (Green)

R01 AG021136 (Tschanz)

R01 HG005092 (Green)

R01 HG06615 (Holm)

K24 AG027841 (Green)

P60 AR047782 (Karlson/Katz)

R01 AG031171 (Qiu)

R01 HG003178 (Wolf)*

P50 HG003170 (Church)

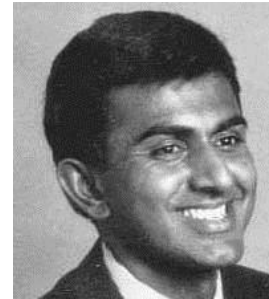
RC1 HG005491 (Holm)*

R21 HG00603 (Wang)

R21 DK084527 (Grant)*

*recently completed

ADNI Data and Publications Committee



ADNI DPC Responsibilities

- Manage access to ADNI data
 - Design and update policies for data access
 - Approve access for each user
 - Maintain table of users and their goals
 - Obtain annual renewal for each user
 - Troubleshoot data access for users
- Manage publication process for ADNI users
 - Review each manuscript for compliance
 - Track publications
 - Intervene with journal editors when necessary

ADNI is a unique experiment in open data access...

credit to Mike Weiner, Arthur Toga and the Informatics Core (Ivani Dos Santos, Karen Crawford), and the Executive Committee

Four Interlocking Policies Create Innovative Data Sharing System

1. Consent for sharing
2. Applications for data access
3. Data downloads
4. Manuscript management

Consent for Sharing

- ... broad sharing explicitly requested
- ... willingness to trust

Consent for Sharing

http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf



Alzheimer's Disease Neuroimaging Initiative (ADNI) Data Sharing and Publication Policy

Introduction

Sharing ADNI data with the general scientific community is an objective established from the onset in the Request for Applications and supported by the ADNI Executive Committee and representatives of the National Institute on Aging. In preparation for this level of data sharing, each of the ADNI sites included the following wording within subject consent documents.

“Study investigators will maintain and be responsible for deciding how your data will be used for future research. All links with your identity will be removed from the data before they are shared. Only de-identified data, which does not include anything that might directly identify you, will be shared with ADNI members and the general scientific community for research purposes.”

The ADNI Executive Committee has created and charged the Data and Publications Committee (DPC) with the task of establishing policies to implement this requirement, as well as recommend procedures for dealing with authorship issues. The DPC began meeting in July, 2005 by teleconference to address these issues. This document outlines the DPC recommendations for sharing data from the ADNI as well as policies for publication and

RORR Statement Locations in ADNI Consents

	ADNI <i>n</i> = 58	ADNI-GO <i>n</i> = 55	ADNI-2 <i>n</i> = 56	PET-PIB <i>n</i> = 15	Total <i>n</i> = 184
Access to research records	2	5	7	-	14
Benefits	7	4	3	8	22
Confidentiality	1	3	4	-	8
Procedures	3	2	2	-	7
Risks	3	5	6	-	14
Sample storage and future use	54	55	55	-	164
Incidental findings	3	2	5	-	10
Use of results in clinical care	1	1	1	-	3
What else do you need to know?	1	-	-	-	1

RORR Statement Referenced in ADNI Consents (ADNI-2)

Type of data referenced in RORR statement	Routinely returned	Not returned	If clinically significant	No statement present	Total
Biomarker	-	53	-	3	56
Genetic	-	55	1	-	56
Imaging, MRI	2	2	7	45	56
Imaging, PET	-	2	1	53	56
Imaging, Florbetapir	-	-	1	55	56
Imaging, Nonspecific	-	-	-	56	56
Lab tests	2	1	-	53	56
Lumbar puncture	2	1	-	53	56
Neuropsychological	-	1	1	54	56
Nonspecific statement	5	4	4	43	56
Total	11	119	15	415	560

Applications for Data Access

- ... simple application process
- ... rapid review of applicants
- ... low threshold for acceptance
- ... no control over analysis plans

ADNI Data Use Agreement

http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Data_Use_Agreement.pdf

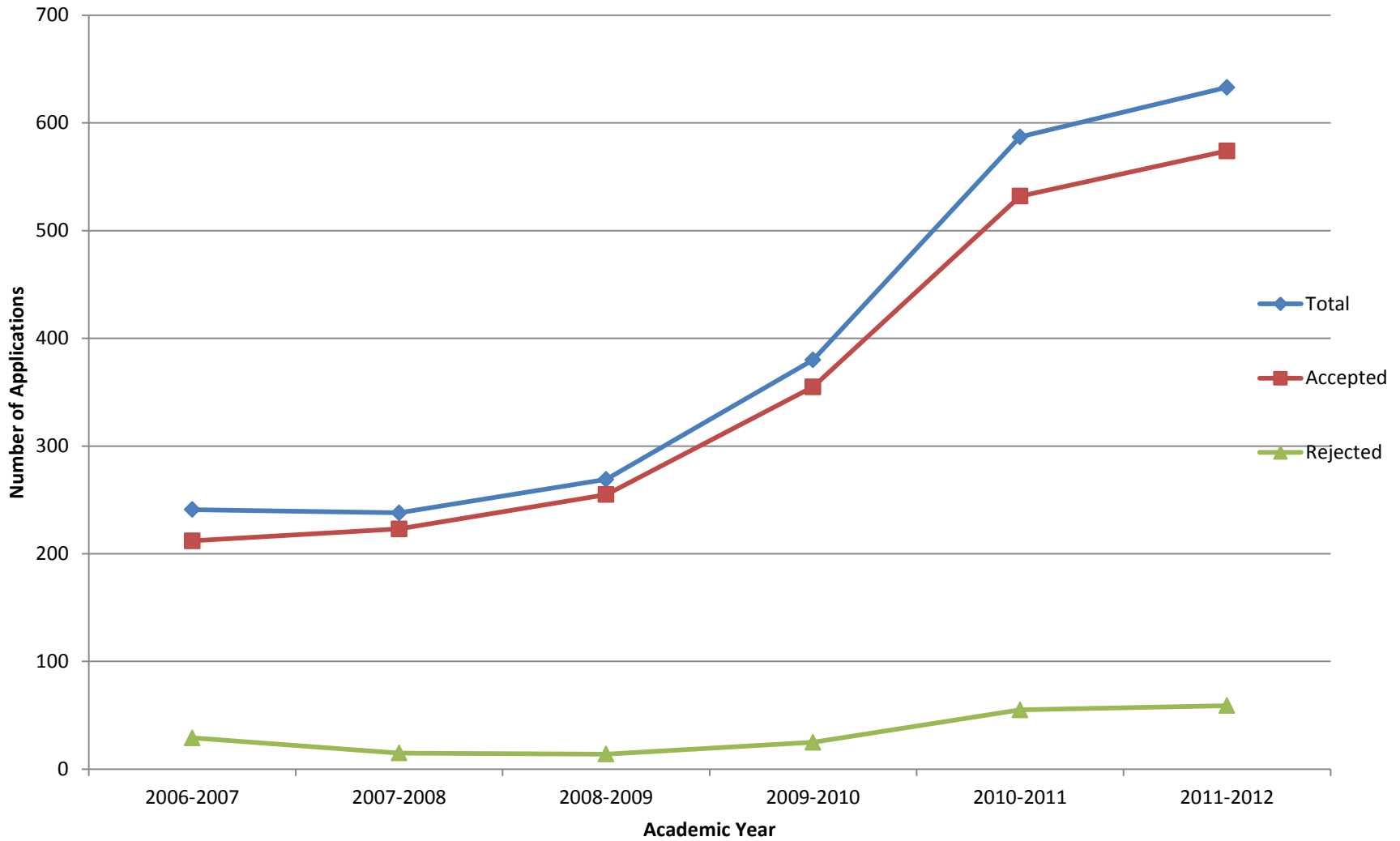


Alzheimer's Disease Neuroimaging Initiative (ADNI) DATA USE AGREEMENT

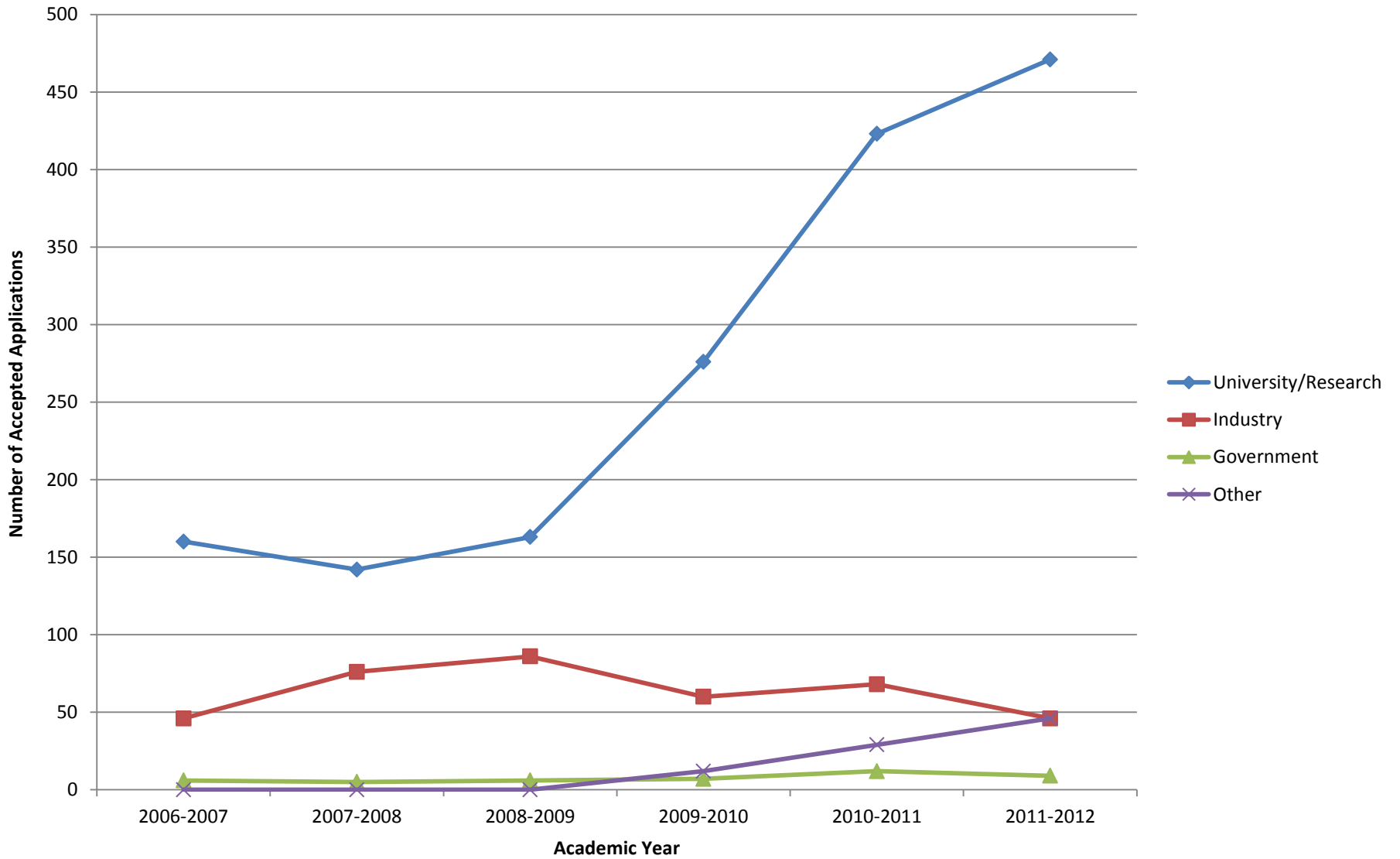
I request access to data collected by the Alzheimer's Disease Neuroimaging Initiative (ADNI) for the purpose of scientific investigation, teaching or the planning of clinical research studies and agree to the following terms.

1. I will receive access to de-identified data and will not attempt to establish the identity of, or attempt to contact any of the ADNI subjects.
2. I will not further disclose these data beyond the uses outlined in this agreement and my data use application and understand that redistribution of data in any manner is prohibited.
3. I will require anyone on my team who utilizes these data, or anyone with whom I share these data to comply with this data use agreement.
4. I will accurately provide the requested information for persons who will use these data and the analyses that are planned using these data.

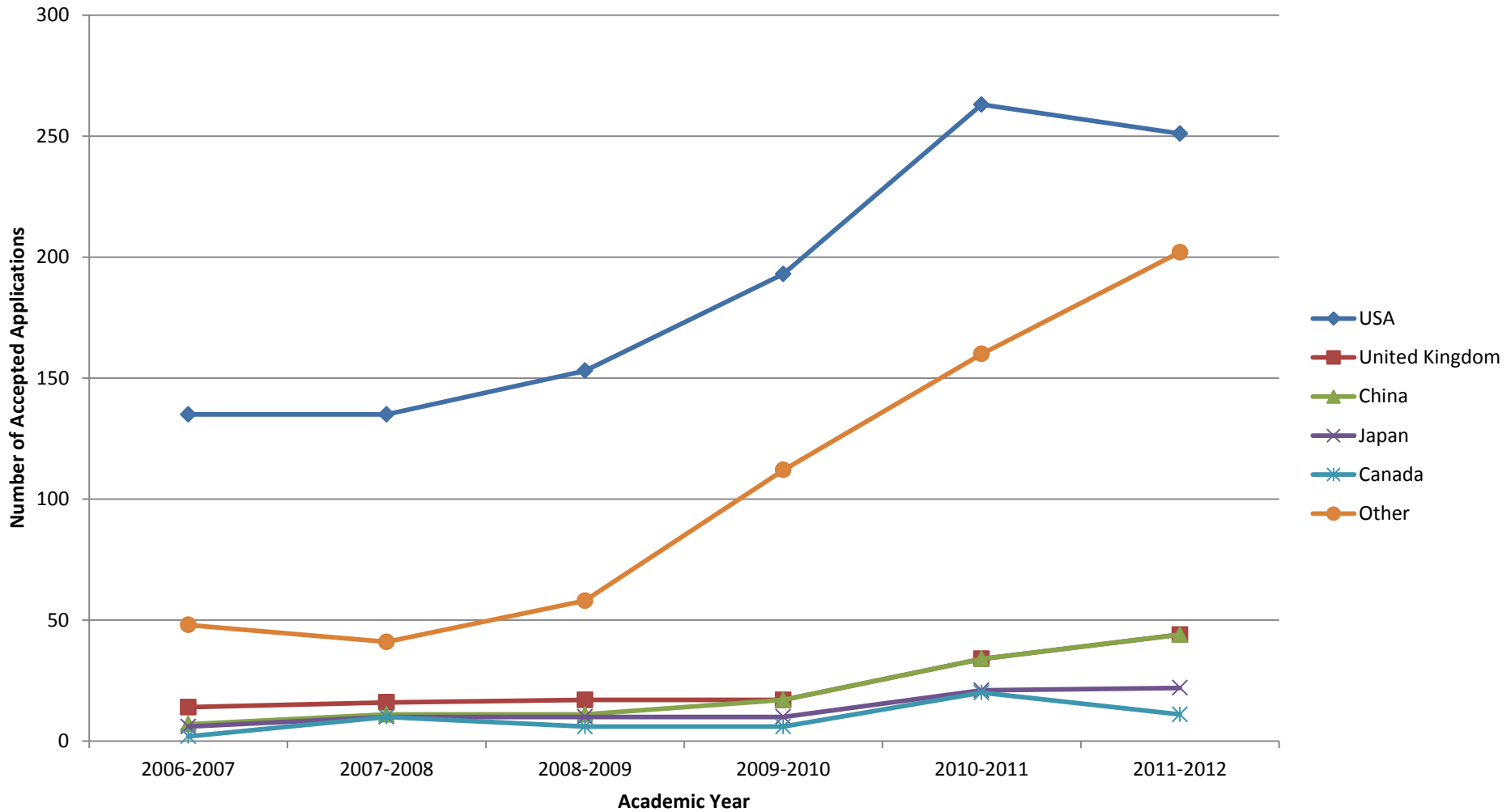
Number of Accepted and Rejected Applications per Year



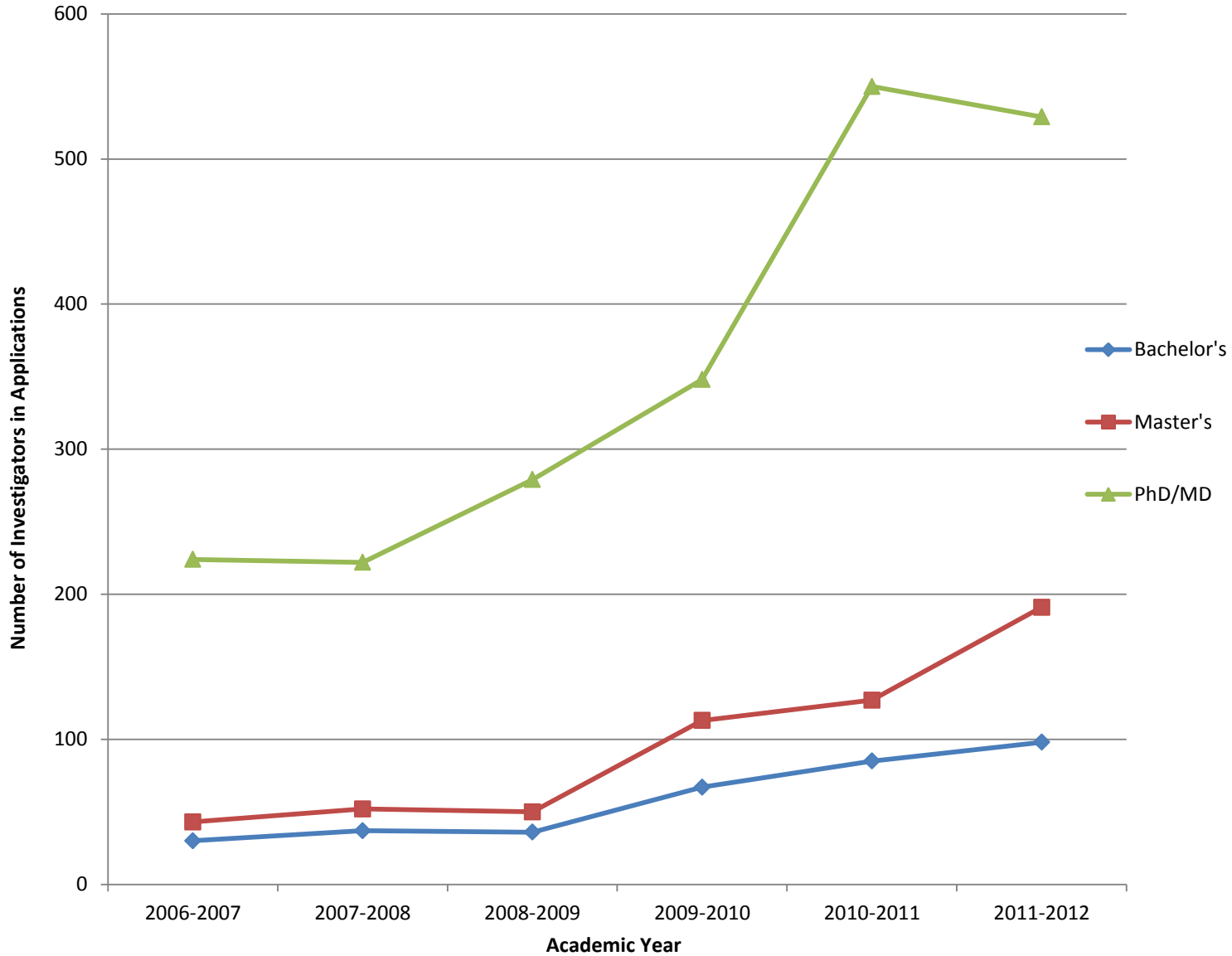
Number of Accepted Applications by Sector per Year



Number of Accepted Applications by Country Per Year



Number of Investigators of Approved Applications by Education Level of User Applicants Per Year



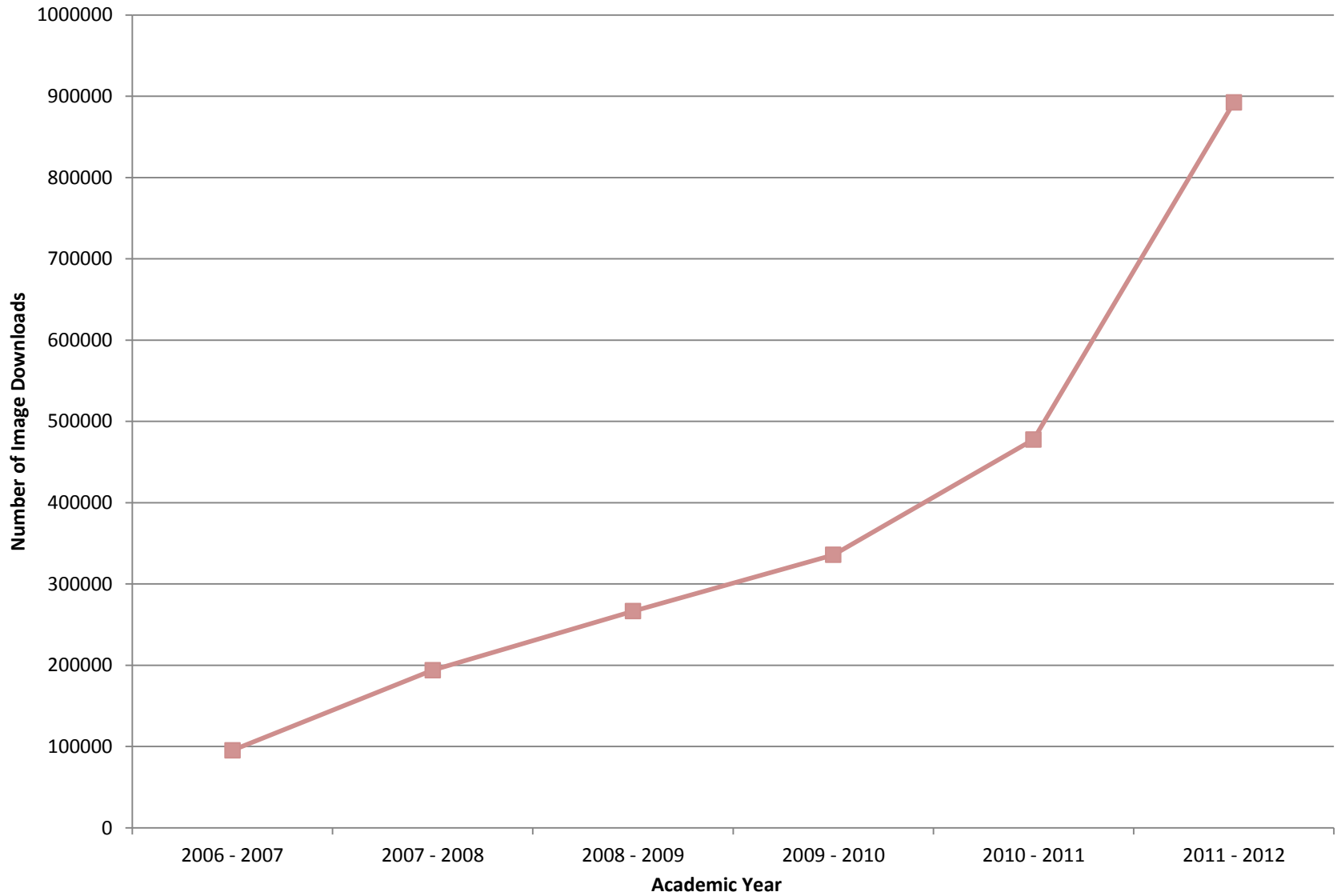
Data Download

- ... available in real time (no embargo)
- ... re-sharing not permitted

Individuals Downloading Data

Number of Users that Downloaded Each File at Least Once	2006	2007	2008	2009	2010	2011
	2007	2008	2009	2010	2011	2012
Imaging	147	189	204	255	356	449
Non-imaging						
Zip files	78	205	266	415	513	223
Individual files	0	0	0	0	224	881

Imaging Downloads Per Year



Manuscript Management

- ... rapid administrative review
- ... limited scientific review
- ... publication tracking

ADNI Publications

<http://adni.loni.ucla.edu/publications/>

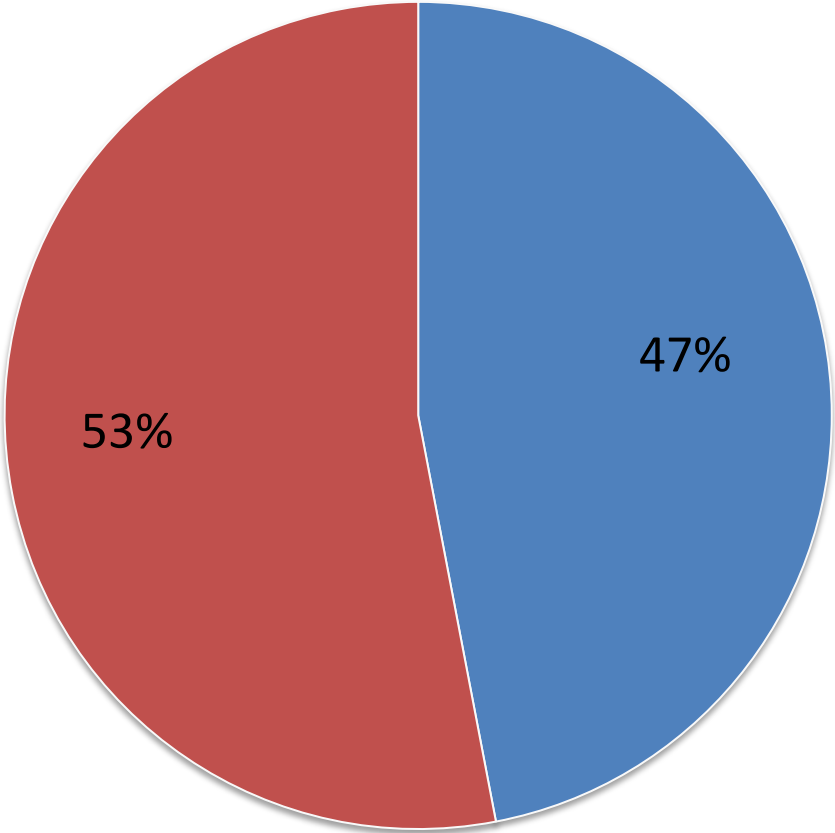
Title	Author(s)	Year	Journal	PMCID/PMID/DOI
Consistent multi-time-point brain atrophy estimation from the boundary shift integral	Leung KK, Ridgway GR, Ourselin S, Fox NC; The Alzheimer's Disease Neuroimaging Initiative	2012	Neuroimage	PMID: 22056457
Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images.	Chu C, Hsu AL, Chou KH, Bandettini P, Lin CP; for the Alzheimer's Disease Neuroimaging Initiative	2012	Neuroimage	PMID: 22166797
Empirical derivation of the reference region for computing diagnostic sensitive (18)fluorodeoxyglucose ratios in Alzheimer's disease based on the ADNI sample.	Rasmussen JM, Lakatos A, van Erp TG, Kruggel F, Keator DB, Fallon JT, Macciardi F, Potkin SG, Alzheimer's Disease Neuroimaging Initiative.	2012	Biochimica et Biophysica acta	PMID: 21958592
Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data.	Cho Y, Seong JK, Jeong Y, Shin SY; for the Alzheimer's Disease Neuroimaging Initiative	2012	Neuroimage	PMID: 22008371
ISOMAP induced manifold embedding and its application to Alzheimer's disease and mild cognitive impairment.	Park H; the ADNI	2012	Neuroscience letters	PMID: 22366258
Pediatric Imaging, Neurocognition, and Genetics Study: Association of common genetic variants in GPCPD1 with scaling of visual cortical surface area in humans.	Bakken TE, Roddey JC, Djurovic S, Akshoomoff N, Amaral DG, Bloss CS, Casey BJ, Chang L, Ernst TM, Gruen JR, Jernigan TL, Kaufmann WE, Kenet T, Kennedy DN, Kuperman JM, Murray SS, Sowell ER, Rimol LM, Mattingsdal M, Melle I, Agartz I, Andreassen OA, Schork NJ, Dale AM; for the Alzheimer's Disease Neuroimaging Initiative	2012	Proceedings of the National Academy of Sciences of the United States of America	PMID: 22343285
Sulcal Span in Alzheimer's Disease, Amnestic Mild Cognitive Impairment, and Healthy Controls.	Reiner P, Jouvent E, Duchesnay E, Cuingnet R, Mangin JF, Chabriat H	2012	Journal of Alzheimer's Disease	PMID: 22297645
Summary Metrics to Assess Alzheimer Disease-Related Hypometabolic Pattern with 18F-FDG PET: Head-to-Head Comparison.	Caroli A, Prestia A, Chen K, Ayutyanont N, Landau SM, Madison CM, Haense C, Herholz K, Nobili F, Reiman EM, Jagust WJ, Frisoni GB; and the EADC-PET Consortium, NEST-DD, and Alzheimer's Disease Neuroimaging Initiative	2012	Journal of Nuclear Medicine	PMID: 22343502
Value of Diagnostic tests to Predict Conversion to Alzheimer's Disease in Young and Old Patients with Amnestic Mild Cognitive Impairment.	Schmand B, Eikelenboom P, van Gool WA	2012	Journal of Alzheimer's Disease	PMID: 22297644

Manuscripts from ADNI

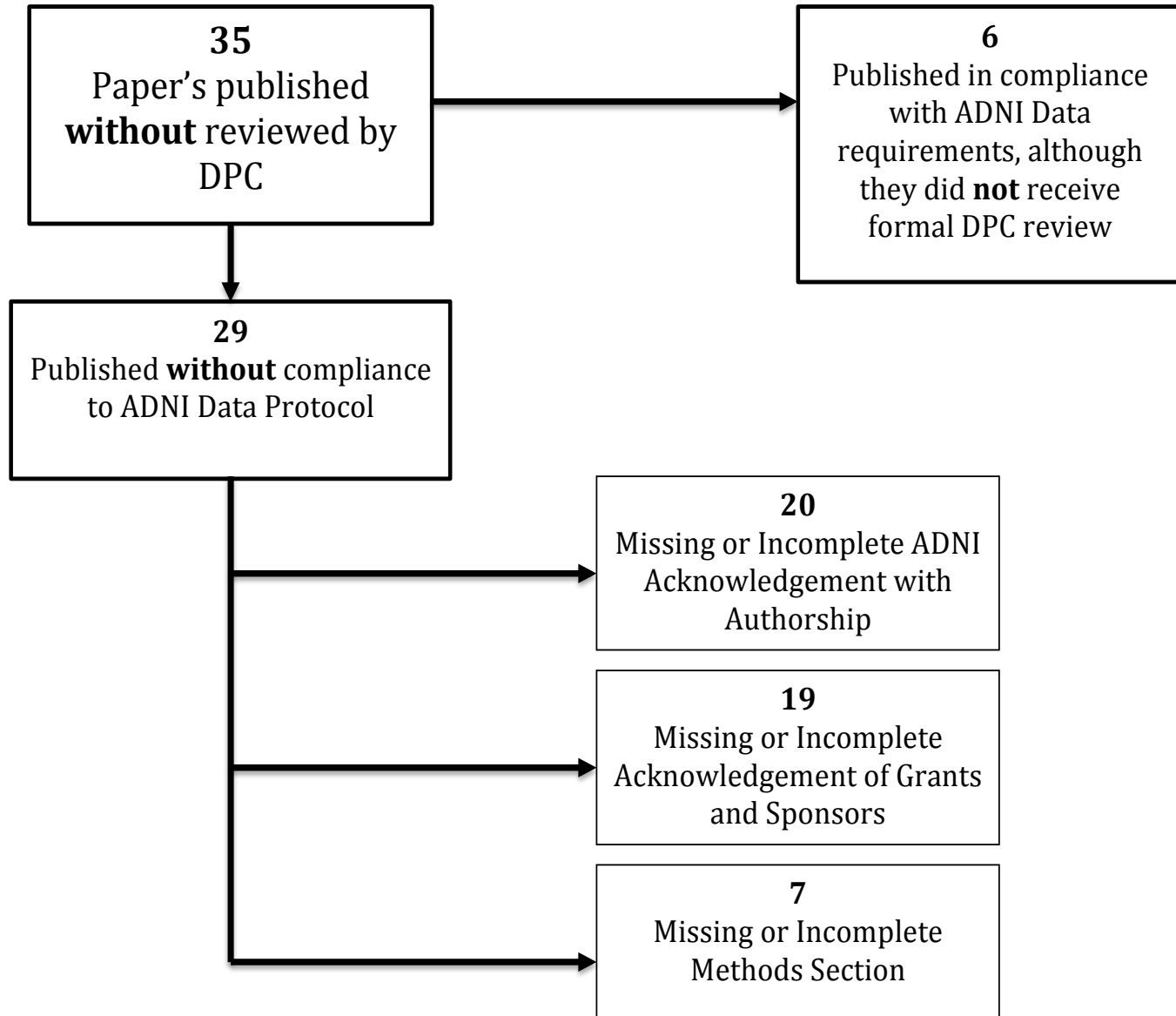
Total Reviewed by ADNI DPC	515
Total Published	247
Total Epubs	24
Total In Press	5
Total Under Revision	2
Total In Submission	220
Total Withdrawn	11
Under Review by ADNI	6
??	??
Total Using ADNI Data Not Reviewed by DPC	35
In compliance with requirements	6
Not in compliance with requirements	29
??	??
Total Manuscripts Using ADNI Data	550

Manuscripts Submitted to DPC Requiring 1 or No Revisions

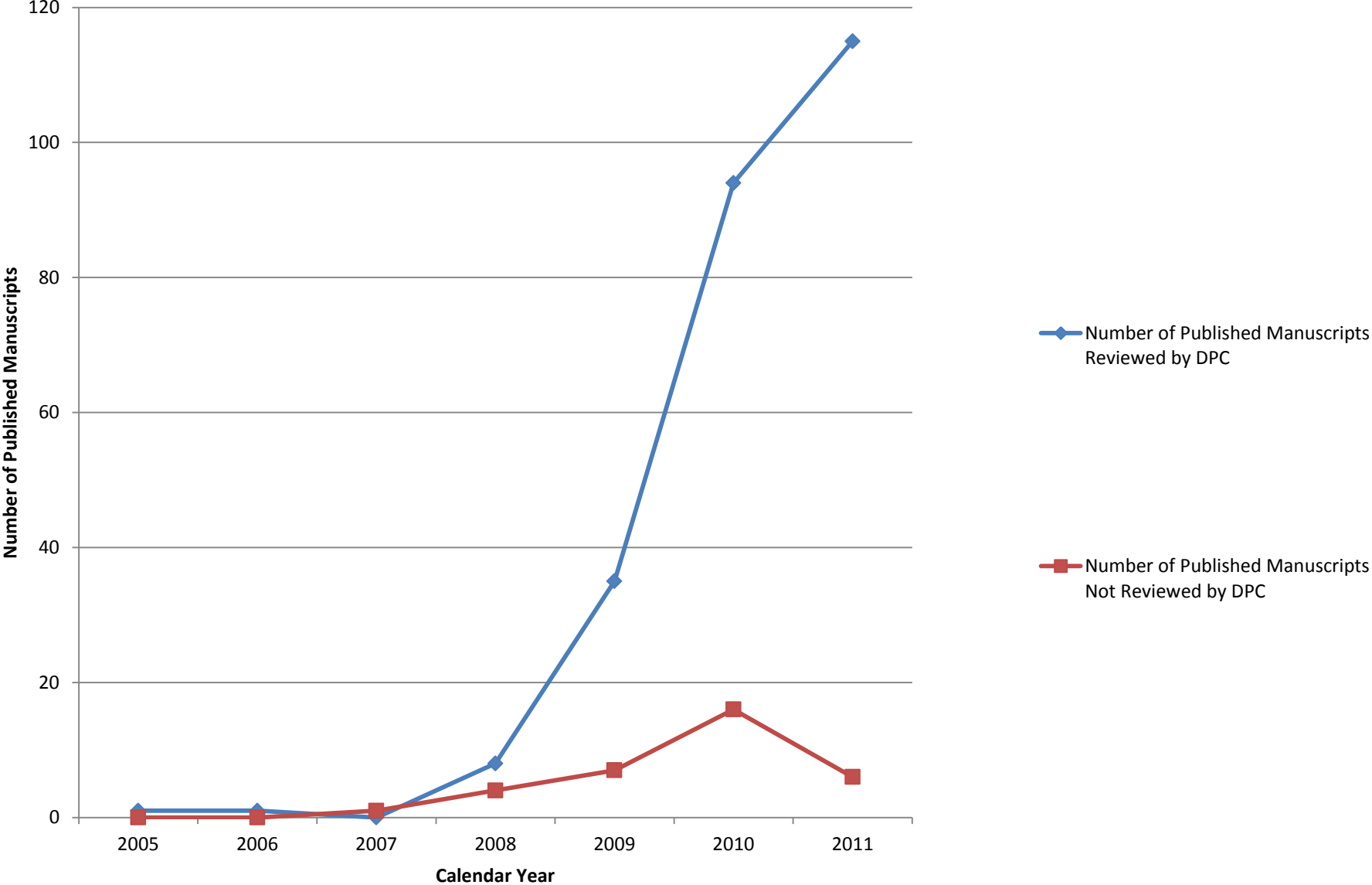
■ No Revision Required ■ 1 Revision Required



Manuscripts Published Without Compliance to ADNI Publishing Policy:



Number of Published Manuscripts with and without DPC Review per Year





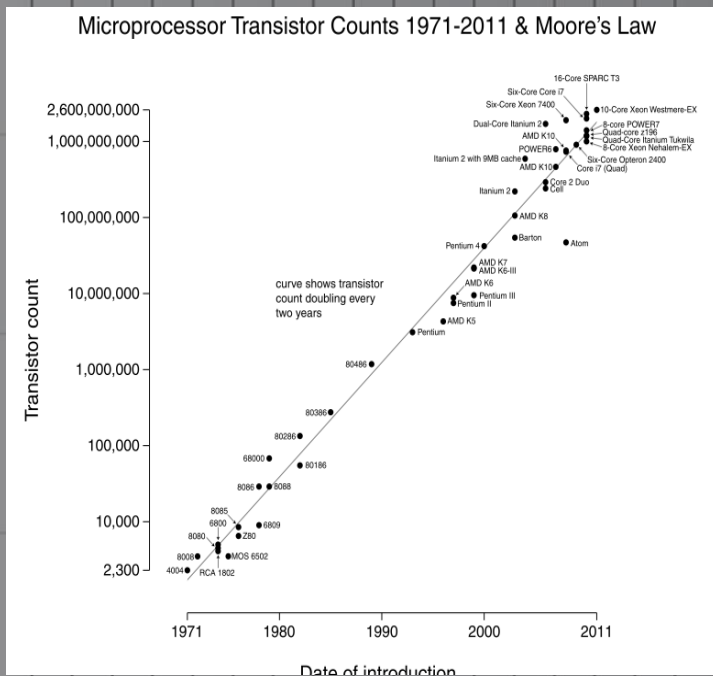
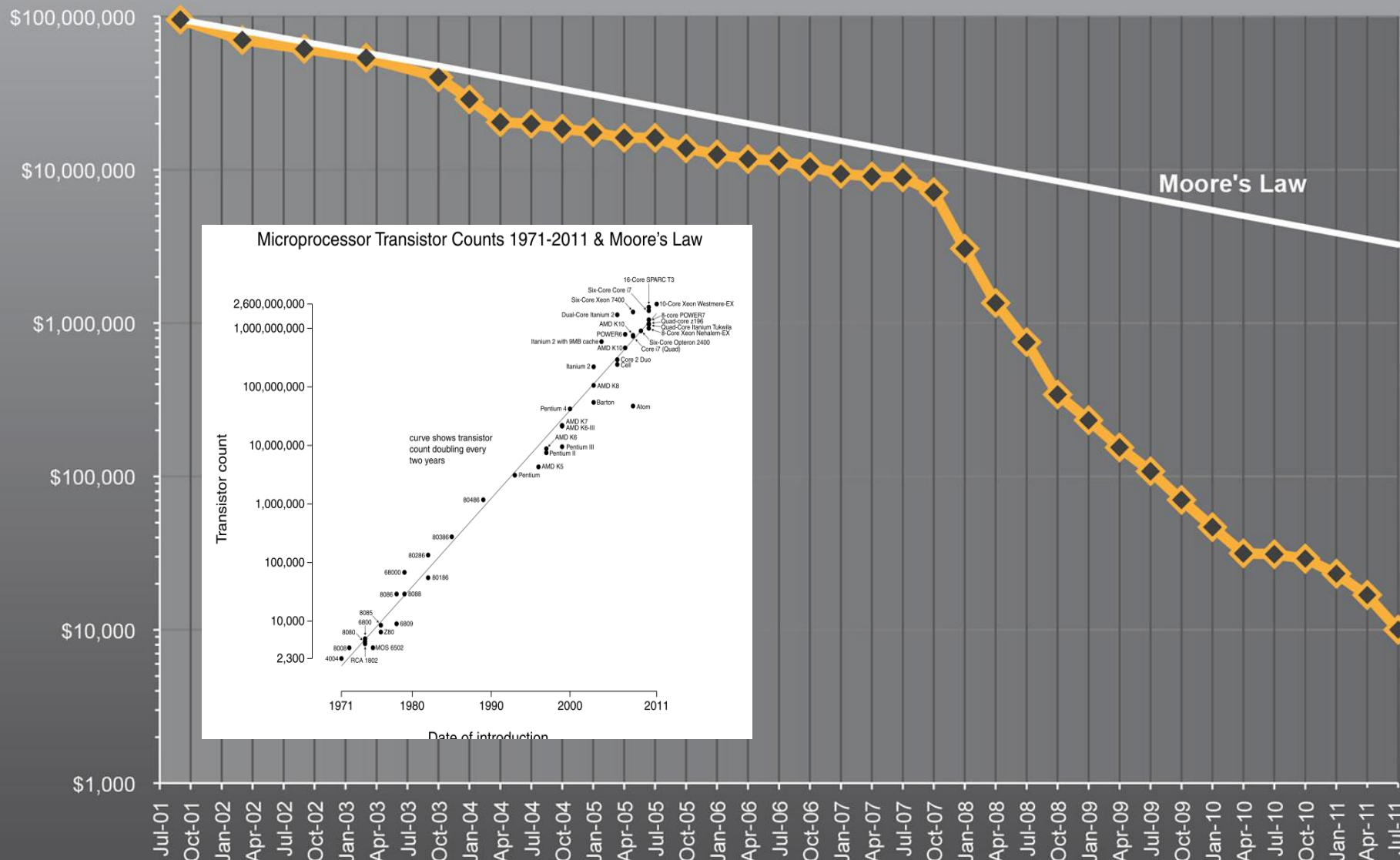
ADNI Sequencing Initiative

Lead: Robert C. Green

Co-Lead: Andy Saykin
Arthur Toga

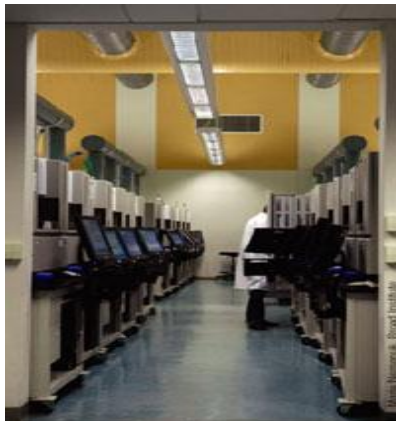
Structure: Genetics Core
Informatics Core
Data & Publications

Cost per Genome



Medical sequencing: a disruptive technology

Time Period	Genomes	Turn-around time	FTEs	Cost per genome
1990-2003	1. NIH reference 2. Celera reference	~5 years	~5,000	~\$2-3 billion
2003-2009	12 additional	~6 months	Dozens	\$300,000→30,000
2010-2014	10^3 - 10^4	2-4 weeks	3-4	\$10,000→1,000
2015-2020	Millions	15 minutes	<<1	\$100



333 machines... each machine generates 1.2 Mb per day

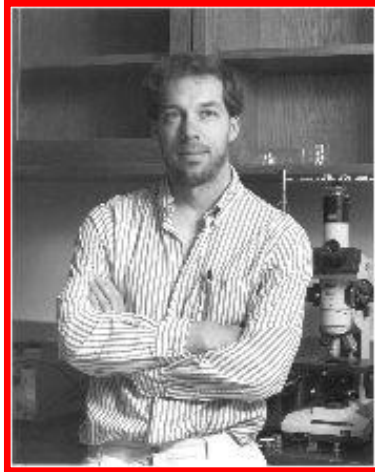
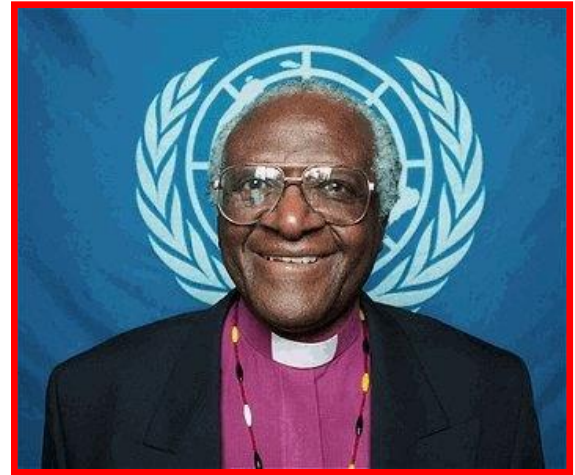
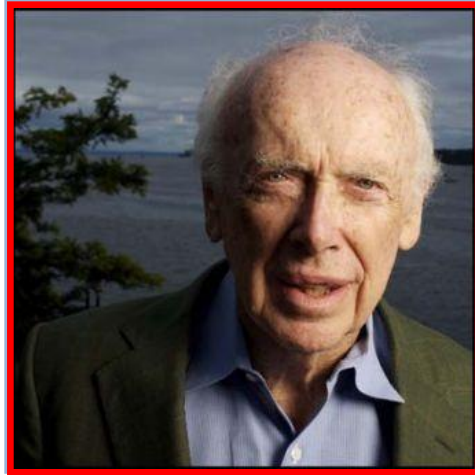
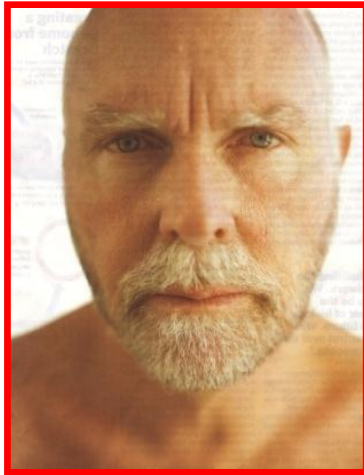


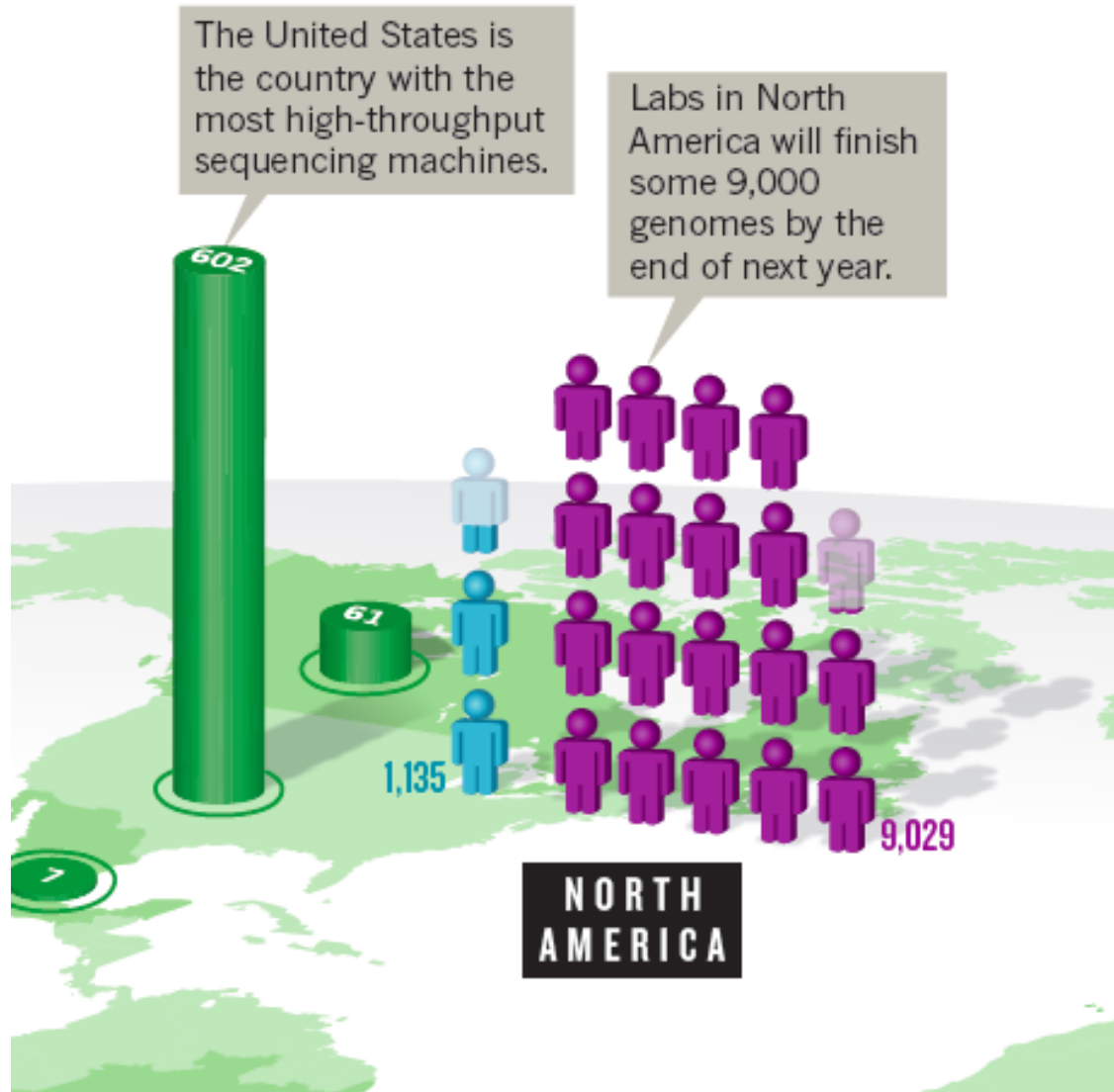
1 machine generates 400 Mb per day



In development
150 Mb in 6 hours

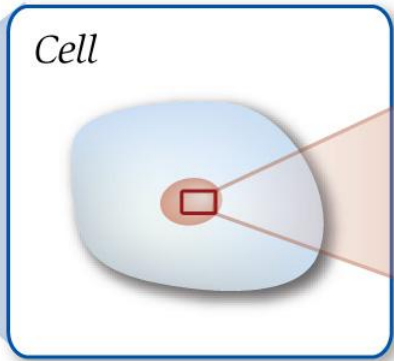
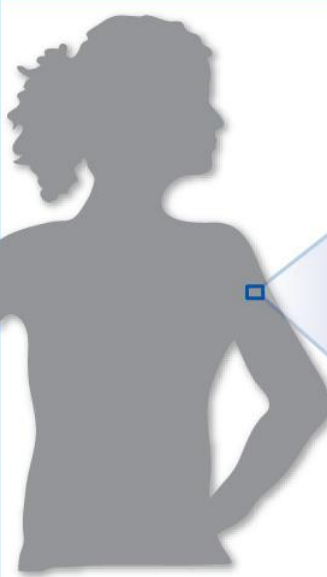
The first humans to be sequenced...





It is estimated that 30,000 genomes will be sequenced by the end of 2011, with over 9000 in the USA. From: Genomes by the thousand. 2010. *Nature* 467.

The human genome contains about 3 billion nucleotides



```
...AGG TTCAGGCATCAGATTCGCAATCGCTTG  
AGCAATCGCTTGCAGATACGAAAGCTTATACC  
TATGTCCTAGGTCAGTGTTTCAAAAAGTTTGT  
TCCATAAAAAGTAACATTGTGCTGCAGGATTT  
CTCAGACGGACCAGTTTGCTAAAGTACTCCGG  
GTGTCTCCACAAAGCTTACATAGAATGTGAAG  
CTTACAAAACATCAGACAAGAGAACATCTC  
CTGGACTGAGTTTAAACACAATTTGGAAA...
```

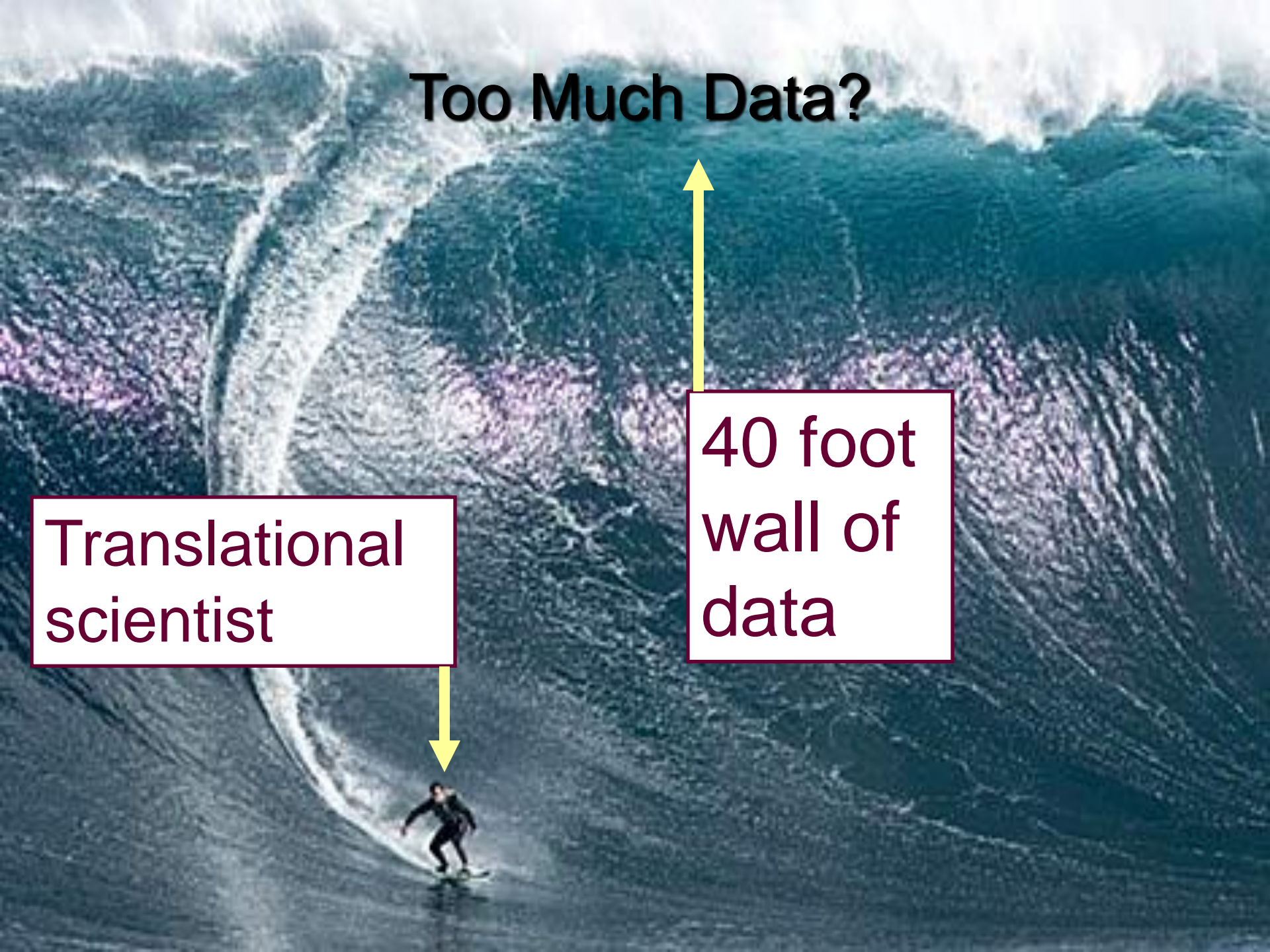
3 billion nucleotides would fill about 200 1,000-page phone books



Too Much Data?

Translational
scientist

40 foot
wall of
data



Too Much Data?

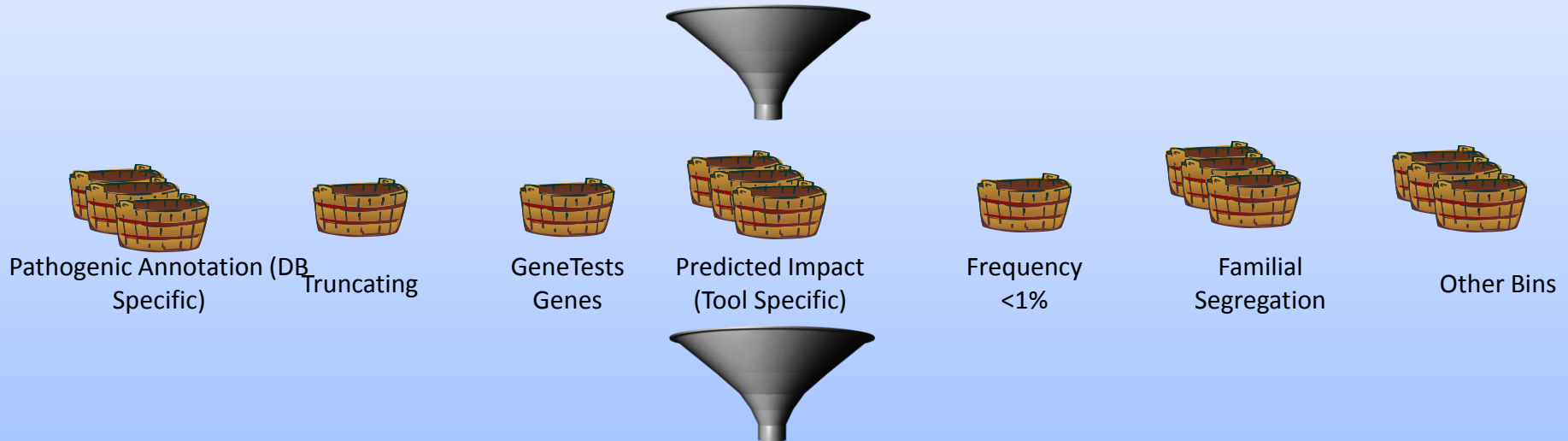


Too Much Data?



Variant Filtration

2 – 5 Million Variants



Variants to Be Manually Assessed

0-2 Causative Variants

ADNI Cases Sent for Sequencing

N = 818

AD ~ 128

MCI ~ 415

Controls ~ 267

Unstable ~ 8

A mutation in *APP* protects against Alzheimer's disease and age-related cognitive decline

Thorlakur Jonsson¹, Jasvinder K. Atwal², Stacy Steinberg¹, Jon Snaedal³, Palmi V. Jonsson^{3,8}, Sigurbjorn Bjornsson³, Hreinn Stefansson¹, Patrick Sulem¹, Daniel Gudbjartsson¹, Janice Maloney², Kwame Hoyte², Amy Gustafson², Yichin Liu², Yanmei Lu², Tushar Bhangale², Robert R. Graham², Johanna Huttenlocher^{1,4}, Gyda Bjornsdottir¹, Ole A. Andreassen⁵, Erik G. Jönsson⁶, Aarno Palotie⁷, Timothy W. Behrens², Olafur T. Magnusson¹, Augustine Kong¹, Unnur Thorsteinsdottir^{1,8}, Ryan J. Watts² & Kari Stefansson^{1,8}

Questions??

