Age-specific signatures of glioblastoma

Serdar Bozdag, Ph.D.

10/21/2013

OPEN ACCESS Freely available online

PLOS ONE

Age-Specific Signatures of Glioblastoma at the Genomic, Genetic, and Epigenetic Levels

Serdar Bozdag[1,2*], Aiguo Li[1], Gregory Riddick[1], Yuri Kotliarov[1,3], Mehmet Baysan[1], Fabio M. Iwamoto[1,4], Margaret C. Cam[1], Svetlana Kotliarova[1], Howard A. Fine[1,5]

1 Neuro-Oncology Branch, National Cancer Institute, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, United States of America, **2** Department of Mathematics, Statistics, and Computer Science, Marquette University, Milwaukee, Wisconsin, United States of America, **3** Center for Human Immunology, Autoimmunity and Inflammation, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **4** The Neurological Institute of New York, College of Physicians and Surgeons, Columbia University, New York, New York, United States of America, **5** New York University Cancer Institute, New York University Langone Medical Center, New York, New York, United States of America

Abstract

Age is a powerful predictor of survival in glioblastoma multiforme (GBM) yet the biological basis for the difference in clinical outcome is mostly unknown. Discovering genes and pathways that would explain age-specific survival difference could generate opportunities for novel therapeutics for GBM. Here we have integrated gene expression, exon expression, microRNA expression, copy number alteration, SNP, whole exome sequence, and DNA methylation data sets of a cohort of GBM patients in The Cancer Genome Atlas (TCGA) project to discover age-specific signatures at the transcriptional, genetic, and epigenetic levels and validated our findings on the REMBRANDT data set. We found major age-specific signatures at all levels including age-specific hypermethylation in polycomb group protein target genes and the upregulation of angiogenesis-related genes in older GBMs. These age-specific differences in GBM, which are independent of molecular subtypes, may in part explain the preferential effects of anti-angiogenic agents in older GBM and pave the way to a better understanding of the unique biology and clinical behavior of older versus younger GBMs.

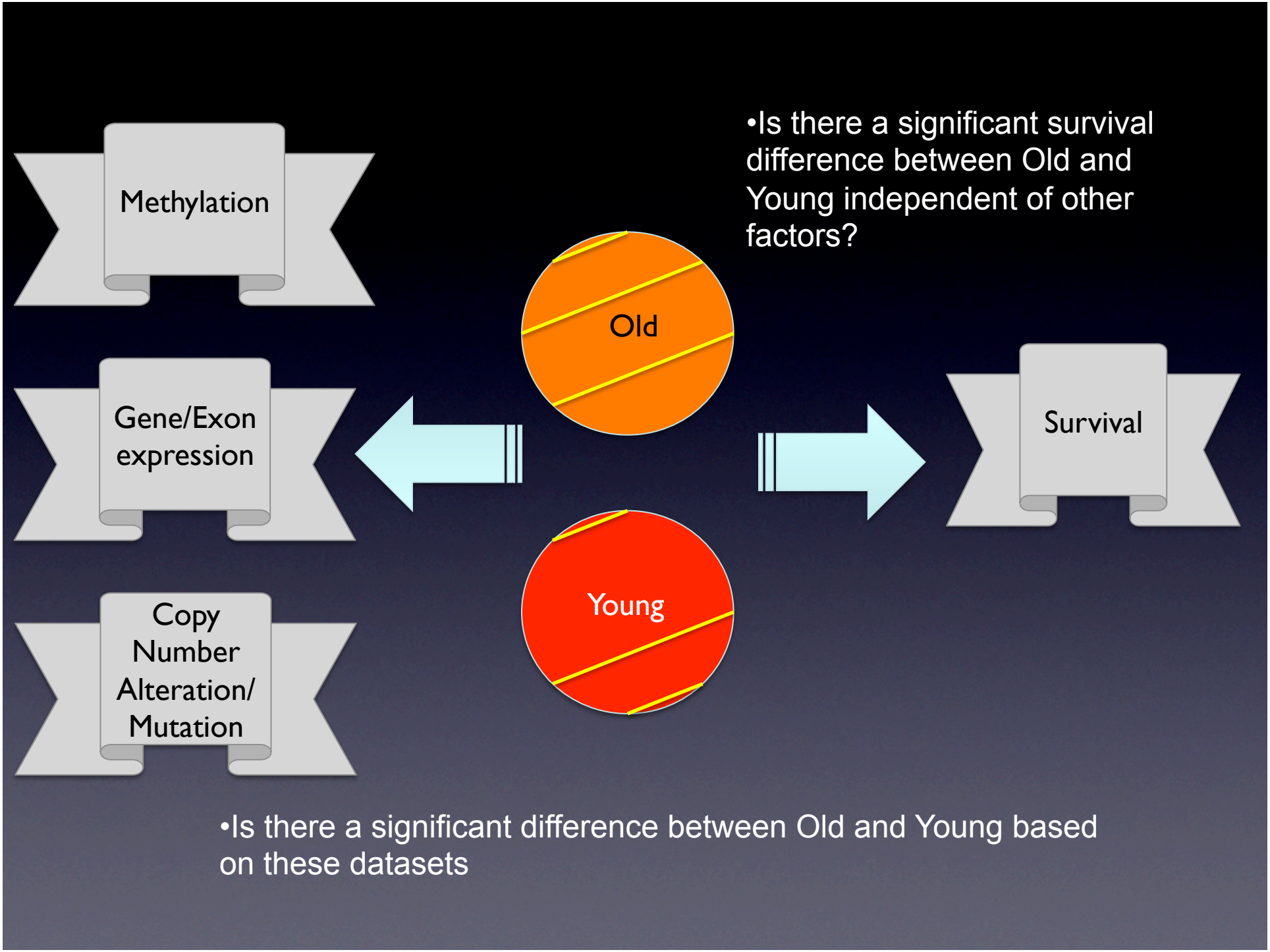
http://www.mscs.mu.edu/~mehdi/seminar/

Motivation

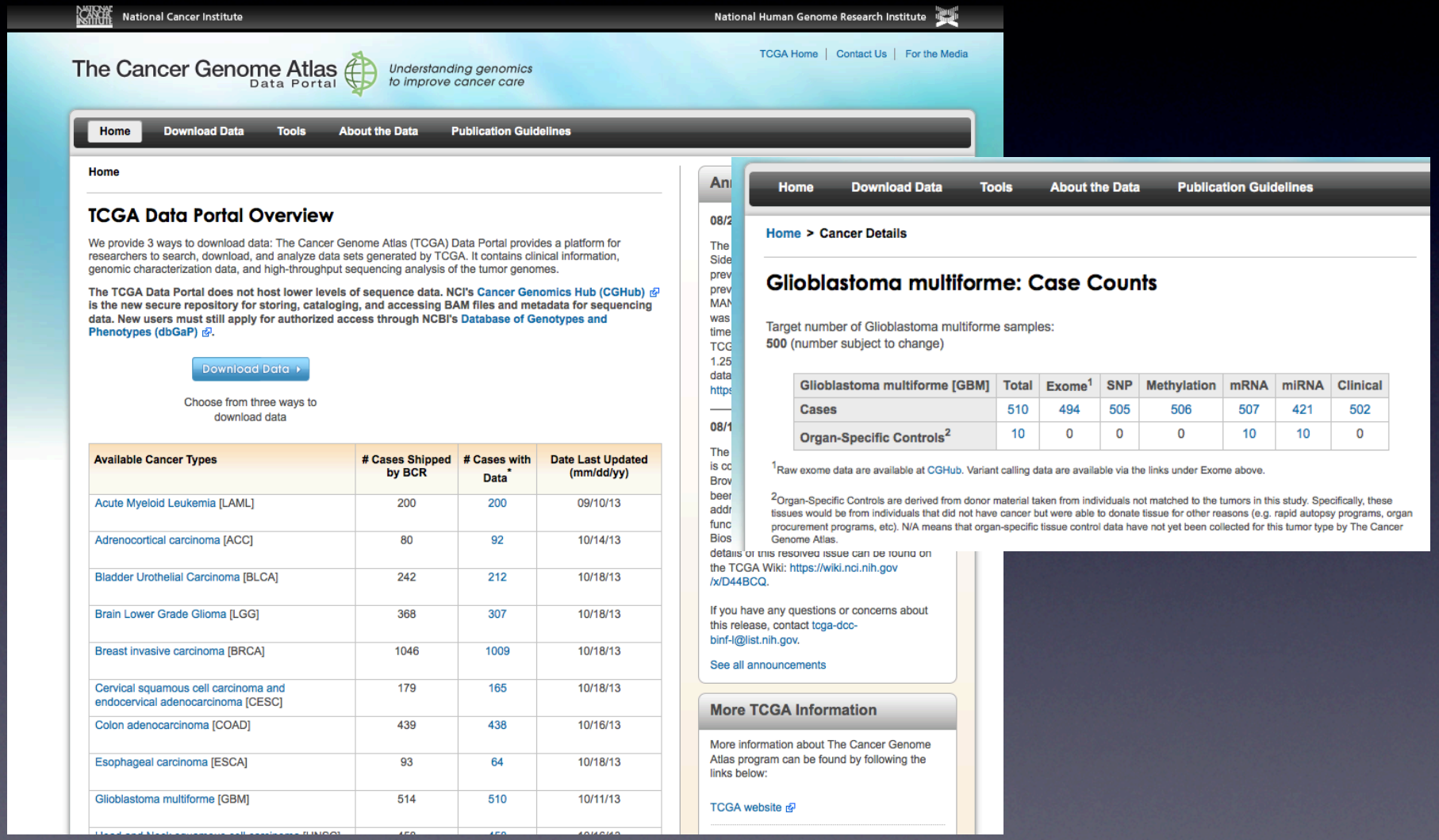
- Glioblastoma multiforme (GBM) is most common malignant type of brain tumor
- GBM patients have a median survival of fourteen months
- Several international projects to generate "big data" to better characterize GBM biology
- It is well known that there is a significant survival difference between old and young GBM patients
- An important remaining questions is "what is the biology behind this survival difference between old and young GBMs"

Objective

- Obtain clinical, genomic, genetic, and epigenetic datasets of GBM patients
  - to verify
    - Age is an independent significant prognostic factor for survival
  - to find
    - Age specific signatures at the genomic, genetic, epigenetic levels


Methylation
Gene/Exon expression
Copy Number Alteration/ Mutation
Old
Young
Survival
•Is there a significant survival difference between Old and Young independent of other factors?
•Is there a significant difference between Old and Young based on these datasets

The Cancer Genome Atlas (TCGA) Project



https://tcga-data.nci.nih.gov/tcga/

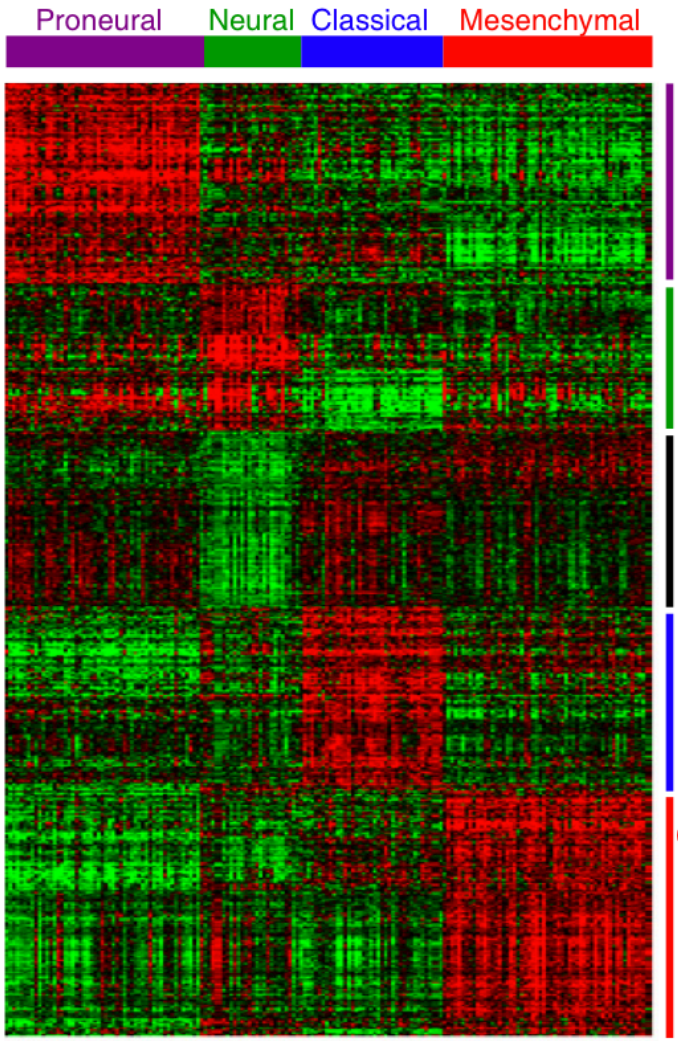
Cell PRESS

Cancer Cell
Article

Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*

Roel G.W. Verhaak,[1,2,17] Katherine A. Hoadley,[3,4,17] Elizabeth Purdom,[7] Victoria Wang,[8] Yuan Qi,[4,5] Matthew D. Wilkerson,[4,5] C. Ryan Miller,[4,6] Li Ding,[9] Todd Golub,[1,10] Jill P. Mesirov,[1] Gabriele Alexe,[1] Michael Lawrence,[1,2] Michael O'Kelly,[1,2] Pablo Tamayo,[1] Barbara A. Weir,[1,2] Stacey Gabriel,[1] Wendy Winckler,[1,2] Supriya Gupta,[1] Lakshmi Jakkula,[11] Heidi S. Feiler,[11] J. Graeme Hodgson,[12] C. David James,[12] Jann N. Sarkaria,[13] Cameron Brennan,[14] Ari Kahn,[15] Paul T. Spellman,[11] Richard K. Wilson,[9] Terence P. Speed,[7,16] Joe W. Gray,[11] Matthew Meyerson,[1,2] Gad Getz,[1] Charles M. Perou,[3,4,8] D. Neil Hayes,[4,5,*] and The Cancer Genome Atlas Research Network

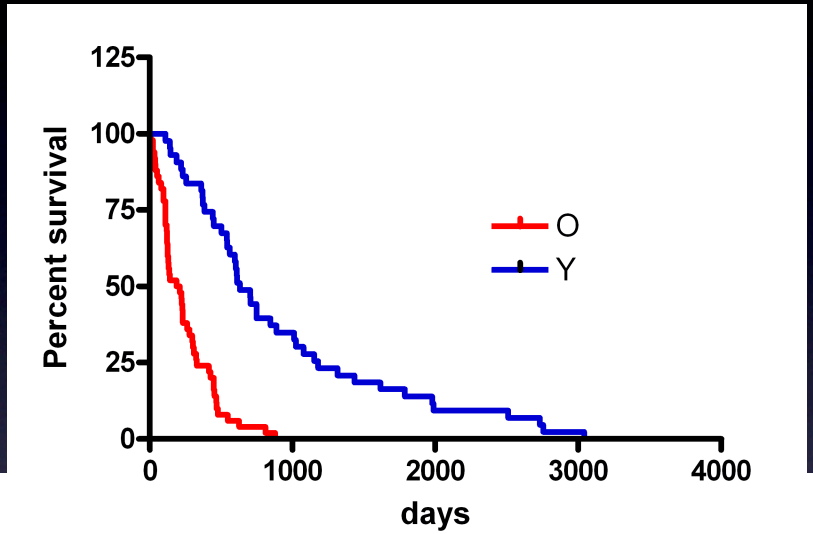
[1]The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142, USA
[2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[3]Department of Genetics
[4]Lineberger Comprehensive Cancer Center

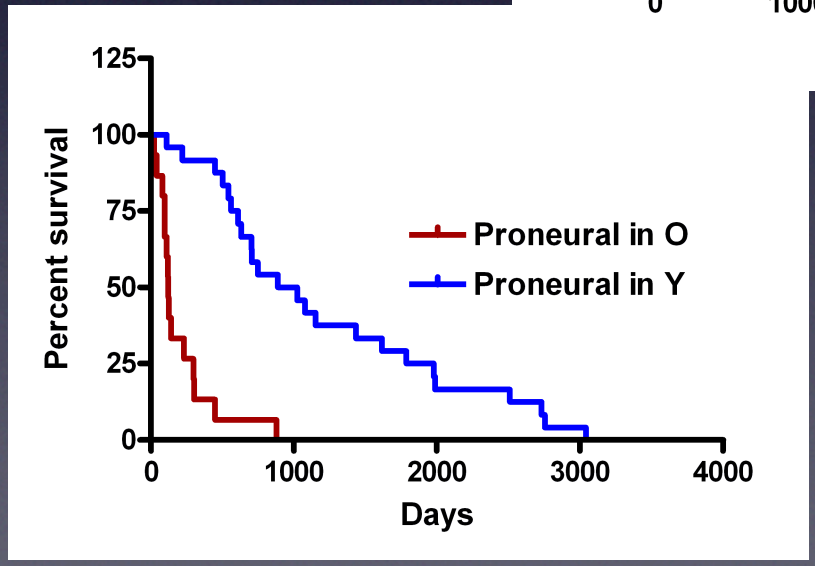


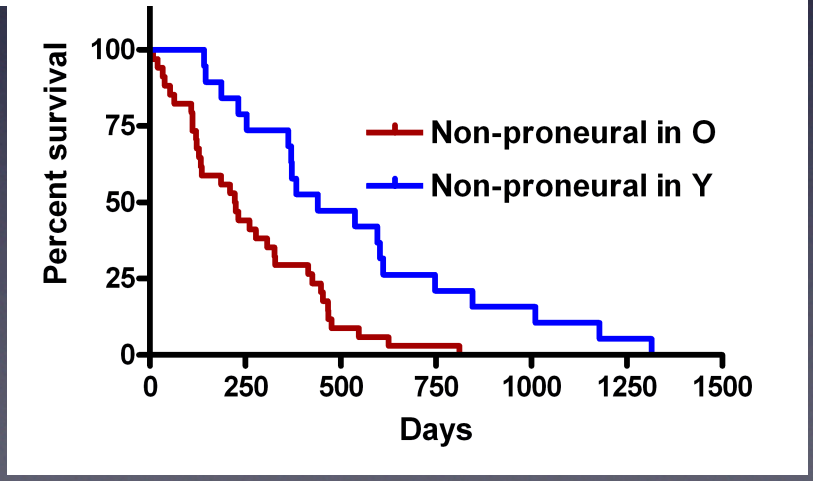
Age trumps molecular subtypes to predict survival

Source	L-R Chi-Square	Prob>ChiSq
Age	36.8861622	**<.0001**
Subtypes	3.13180449	0.3717
Subtypes*Age	1.57077625	0.666

Age trumps molecular subtypes to predict survival







Cancer Cell
Article

Cancer Cell *17*, 1–13, May 18, 2010 ©2010 Elsevier Inc. 1

Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma

Houtan Noushmehr,[1,13] Daniel J. Weisenberger,[1,13] Kristin Diefes,[2,13] Heidi S. Phillips,[3] Kanan Pujara,[3] Benjamin P. Berman,[1] Fei Pan,[1] Christopher E. Pelloski,[4] Erik P. Sulman,[4] Krishna P. Bhat,[2] Roel G.W. Verhaak,[5,6] Katherine A. Hoadley,[7,8] D. Neil Hayes,[7,8] Charles M. Perou,[7,8] Heather K. Schmidt,[9] Li Ding,[9] Richard K. Wilson,[9] David Van Den Berg,[1] Hui Shen,[1] Henrik Bengtsson,[10] Pierre Neuvial,[10] Leslie M. Cope,[11] Jonathan Buckley,[1,12] James G. Herman,[11] Stephen B. Baylin,[11] Peter W. Laird,[1,14,*] Kenneth Aldape,[2,14] and The Cancer Genome Atlas Research Network

SUMMARY

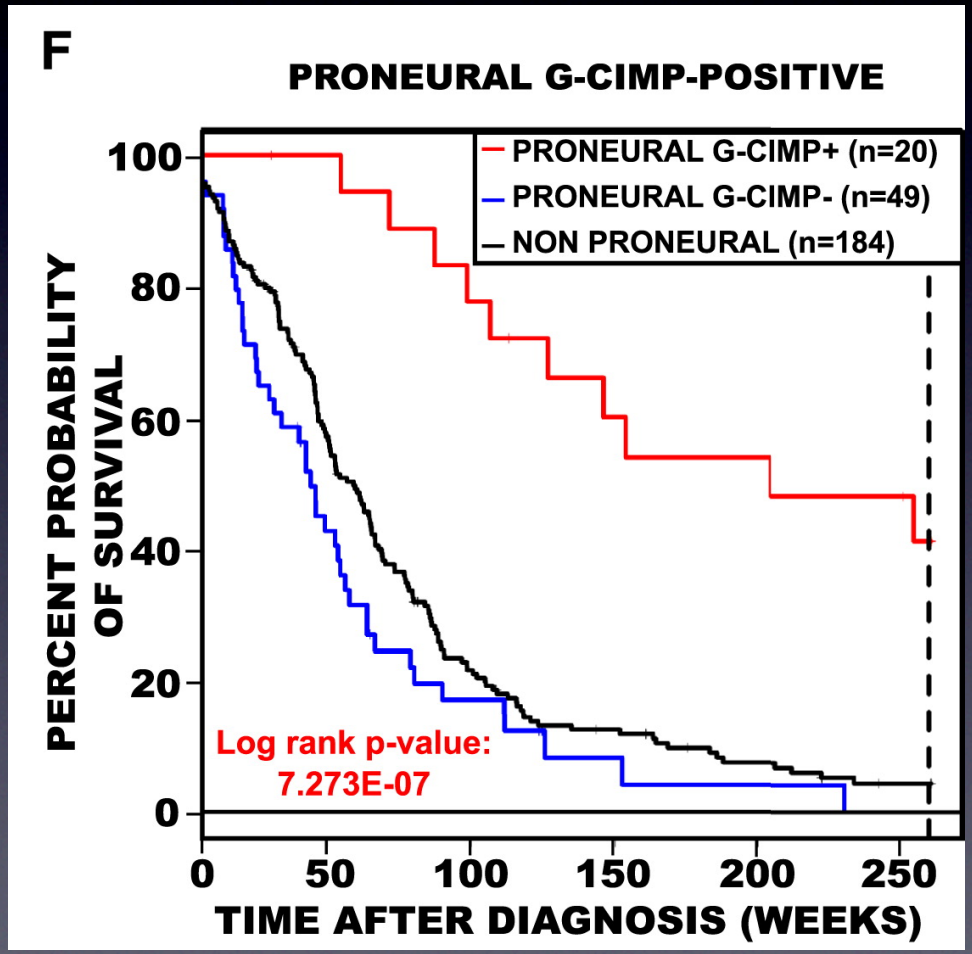
We have profiled promoter DNA methylation alterations in 272 glioblastoma tumors in the context of The Cancer Genome Atlas (TCGA). We found that a distinct subset of samples displays concerted hypermethylation at a large number of loci, indicating the existence of a glioma-CpG island methylator phenotype (G-CIMP). We validated G-CIMP in a set of non-TCGA glioblastomas and low-grade gliomas. G-CIMP tumors belong to the proneural subgroup, are more prevalent among lower-grade gliomas, display distinct copy-number alterations, and are tightly associated with *IDH1* somatic mutations. Patients with G-CIMP tumors are younger at the time of diagnosis and experience significantly improved outcome. These findings identify G-CIMP as a distinct subset of human gliomas on molecular and clinical grounds.

Age is an independent predictor of survival

F	Hazard Ratio	P-Value
Tumor grade	2.1	<0.000001
Patient Age (in decades)	1.2	0.000026
G-CIMP status	0.4	0.000020

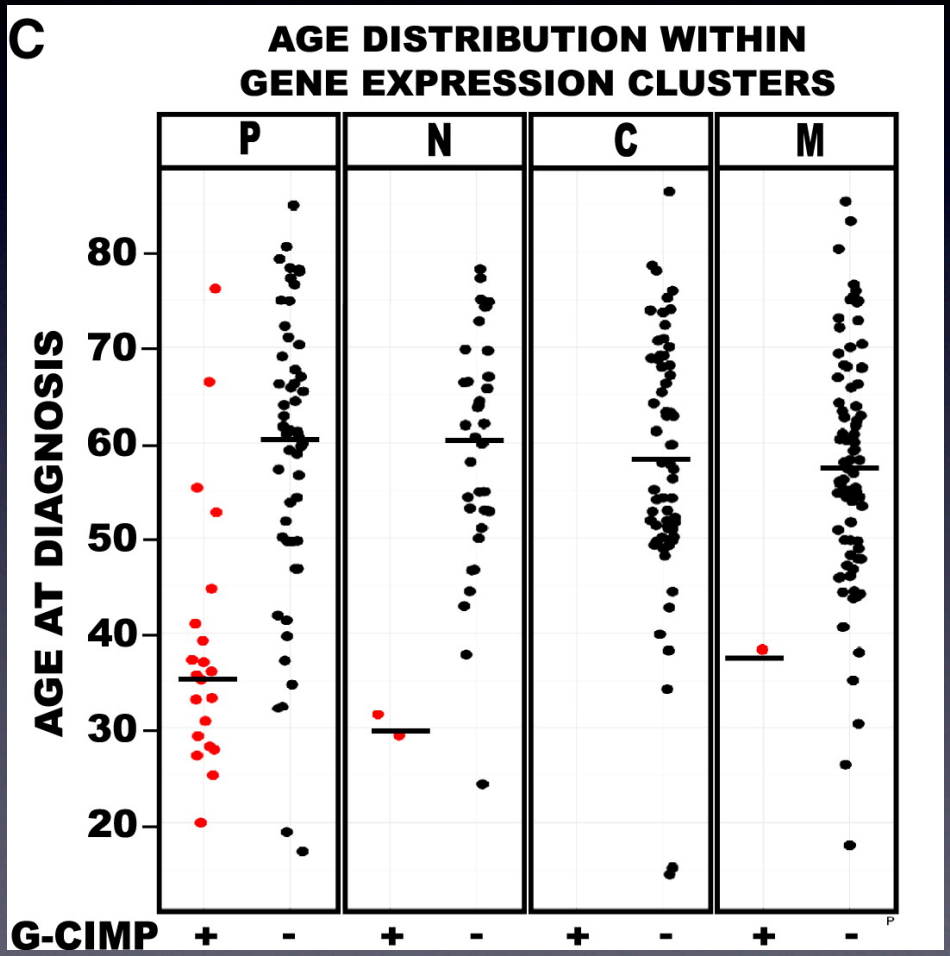
Noushmehr, et al. Cancer Cell 2010

G-CIMP+ proneurals and G-CIMP- proneurals have different biology



Noushmehr, et al. Cancer Cell 2010

G-CIMP+ samples are younger than G-CIMP- samples

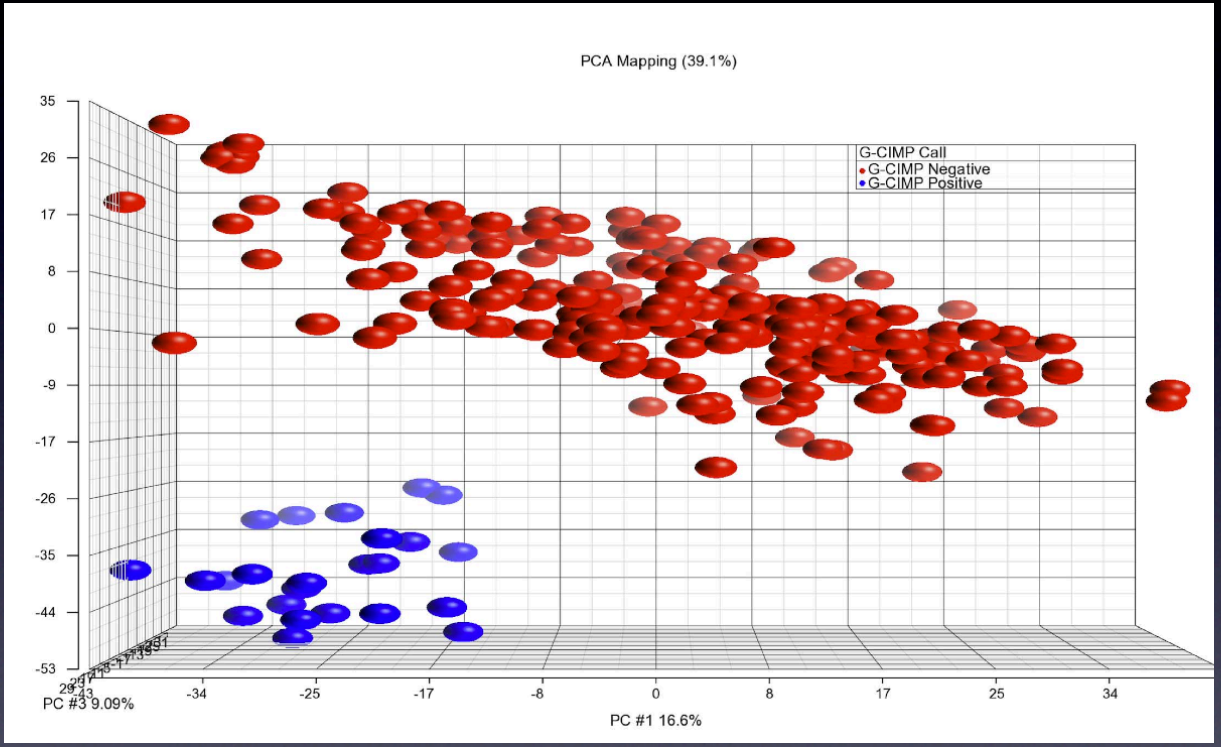


Noushmehr, et al. Cancer Cell 2010

We need to remove G-CIMP effect from sample set

- We would like to find age-specific changes at expression/methylation
- However, G-CIMP status affect changes at expression/ methylation level
- We remove G-CIMP+ samples

G-CIMP+ vs. G-CIMP-
(PCA on methylation data)



- 281 samples in methylation data
- Probesets std. dev >= 0.2 selected

G-CIMP prediction from gene expression

- Expression profiles of samples with known G-CIMP status are used as training
- K-nearest algorithm with cross-validation applied
- Two gene expression datasets were used for prediction
- Consensus G-CIMP calls are saved

OPEN ACCESS Freely available online

PLOS ONE

G-Cimp Status Prediction Of Glioblastoma Samples Using mRNA Expression Data

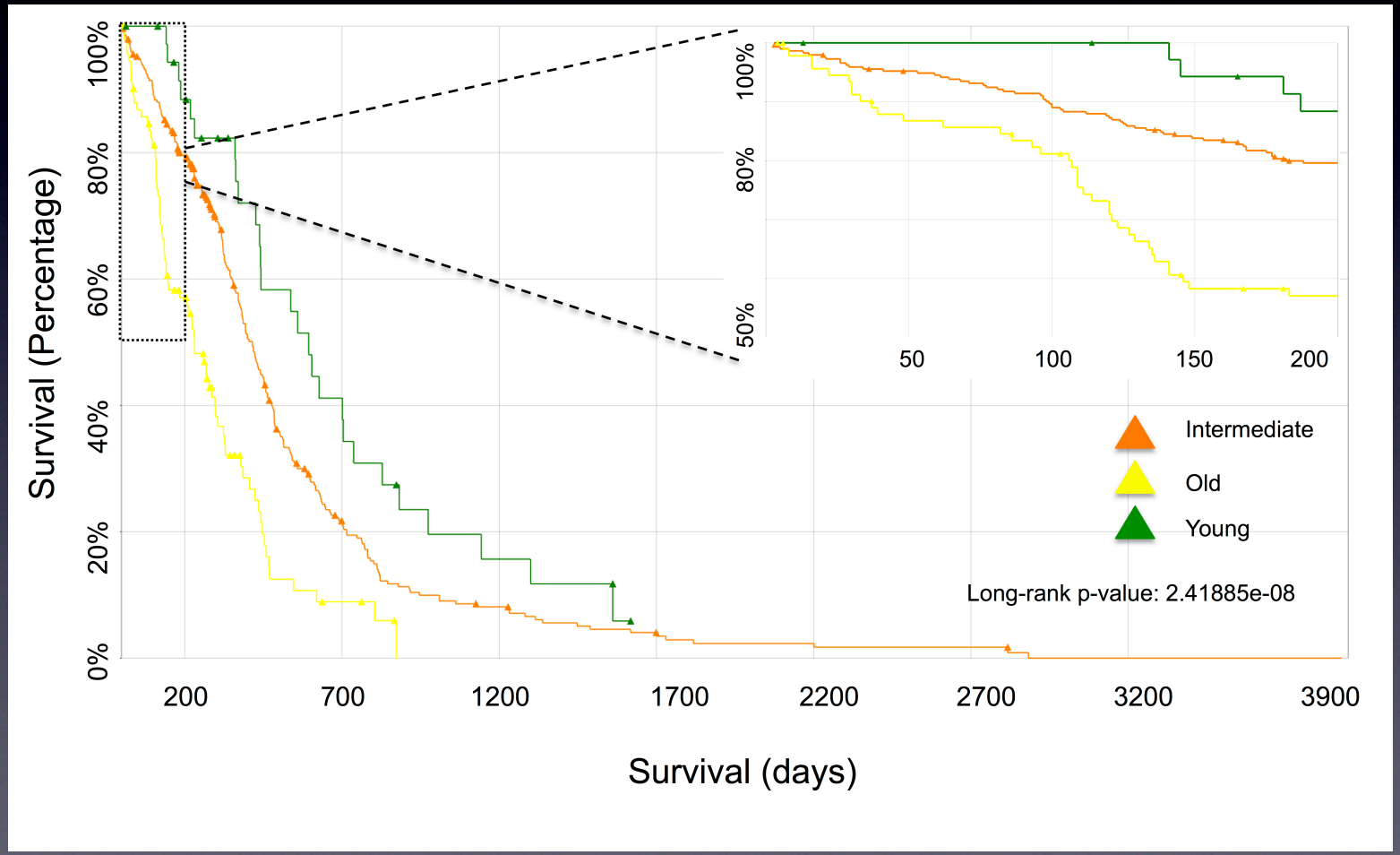
Mehmet Baysan[1], Serdar Bozdag[2], Margaret C. Cam[1], Svetlana Kotliarova[1], Susie Ahn[1], Jennifer Walling[1], Jonathan K. Killian[3], Holly Stevenson[3], Paul Meltzer[3], Howard A. Fine[1,4]*

1 Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **2** Department of Mathematics, Statistics, and Computer Science, Marquette University, Milwaukee, Wisconsin, United States of America, **3** Genetics Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **4** New York University Cancer Institute, New York University Langone Medical Center, New York, New York, United States of America

Abstract

Glioblastoma Multiforme (GBM) is a tumor with high mortality and no known cure. The dramatic molecular and clinical heterogeneity seen in this tumor has led to attempts to define genetically similar subgroups of GBM with the hope of developing tumor specific therapies targeted to the unique biology within each of these subgroups. Recently, a subset of relatively favorable prognosis GBMs has been identified. These glioma CpG island methylator phenotype, or G-CIMP tumors, have distinct genomic copy number aberrations, DNA methylation patterns, and (mRNA) expression profiles compared to other GBMs. While the standard method for identifying G-CIMP tumors is based on genome-wide DNA methylation data, such data is often not available compared to the more widely available gene expression data. In this study, we have developed and evaluated a method to predict the G-CIMP status of GBM samples based solely on gene expression data.

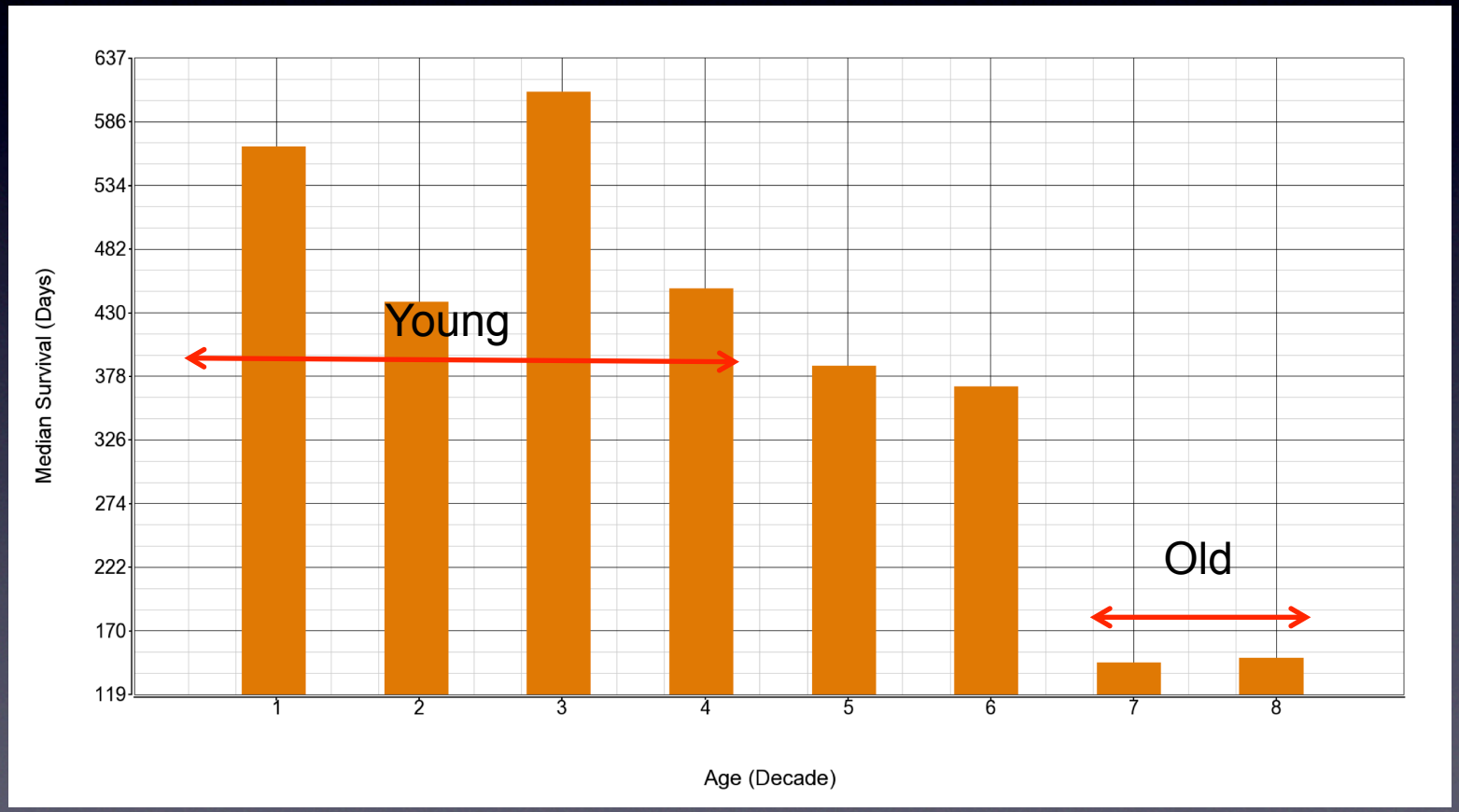
Age is significant factor of survival within G-CIMP- samples



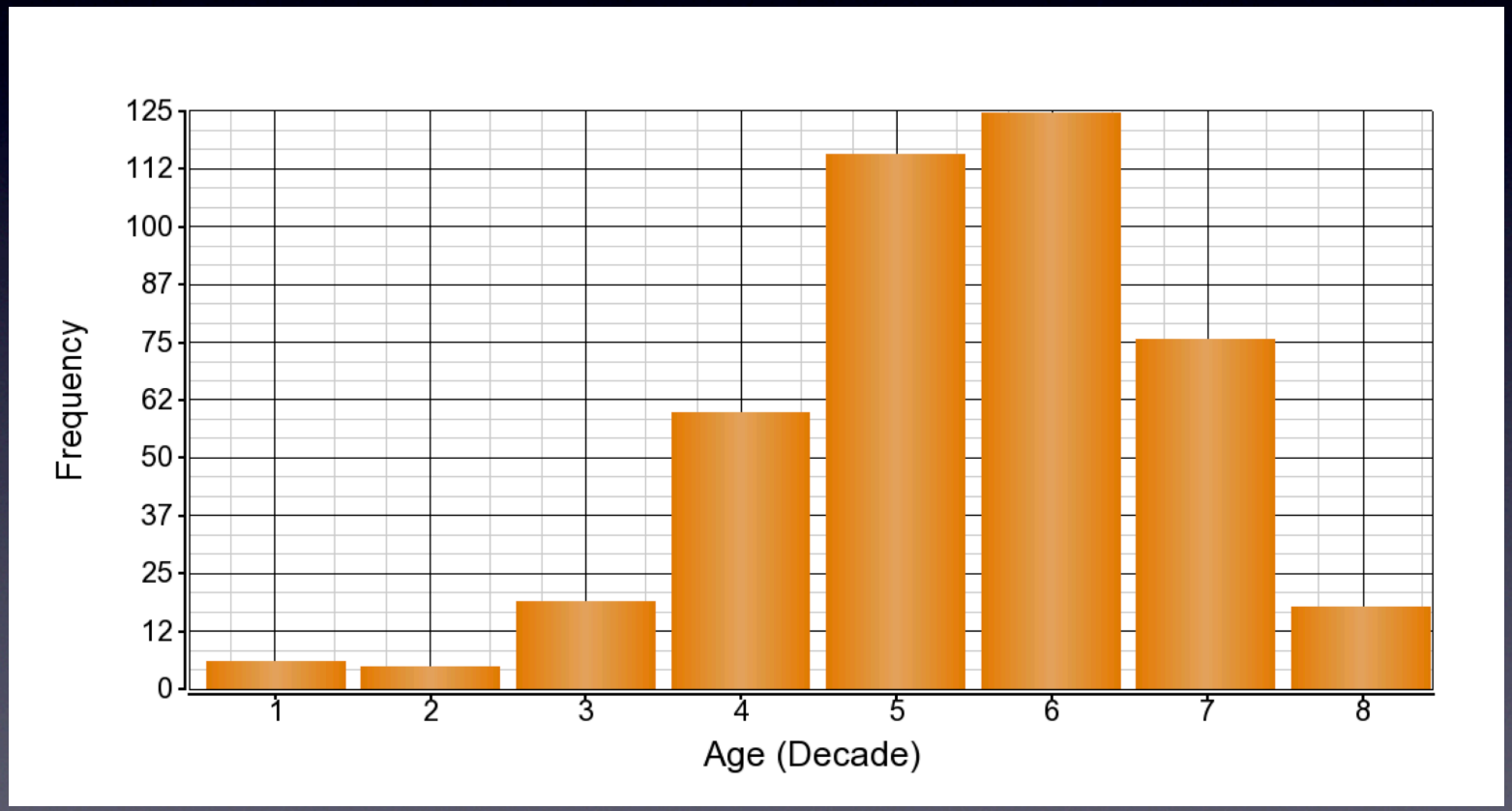
Old vs. Young

- Gene expression
  - Agilent
  - Affymetrix
- Exon expression
- Methylation
- CNA
- Mutation

What is the definition of "old" and "young"



Number of sample per age group



Data set

Data Type	Platform	Level[1]	Institute	# Old[2]	# Young[2]	Total
Gene expression	Affymetrix HT Human Genome U133 Array Plate Set	2	Broad Institute of MIT and Harvard	92	37	422
Exon expression	Affymetrix Human Exon 1.0 ST Array	3	Lawrence Berkeley National Laboratory	80	34	382
Gene expression	Agilent 244K Custom Gene Expression G4502A	2	University of North Carolina	92	37	420
miRNA expression	Agilent 8×15K Human miRNA-specific microarray	3	University of North Carolina	80	34	385
Methylation	Illumina Infinium Human DNA Methylation 27	2	Johns Hopkins/University of Southern California	56	22	256
Copy Number	Agilent Human Genome CGH Microarray 244A	3	Memorial Sloan-Kettering Cancer Center	87	36	406
SNP	Affymetrix Genome-Wide Human SNP Array 6.0	3	Broad Institute of MIT and Harvard	88	32	390
SNP	Illumina 550K Infinium HumanHap550 SNP Chip	3	HudsonAlpha Institute for Biotechnology	78	33	376
Whole Exome Sequence	Illumina Genome Analyzer IIx	N/A	Broad Institute of MIT and Harvard	55	12	202

[1]Level 2 refers to probeset-level data and level 3 refers to gene-level data for expression and methylation data sets. Level 3 refers to segmented data for copy number and SNP data sets. There is no level number for whole exome sequence data set as we just used the mutations derived from this data set.
[2]Old and Young refer to samples ≥70 and ≤40 years old, respectively.

Computing age-specific significant differentially expressed genes (DEGs)

- Two methods have been applied
  - Two sample t-test (old vs. young)
  - Linear regression (SAM) where age is a continuous variable

DEGs

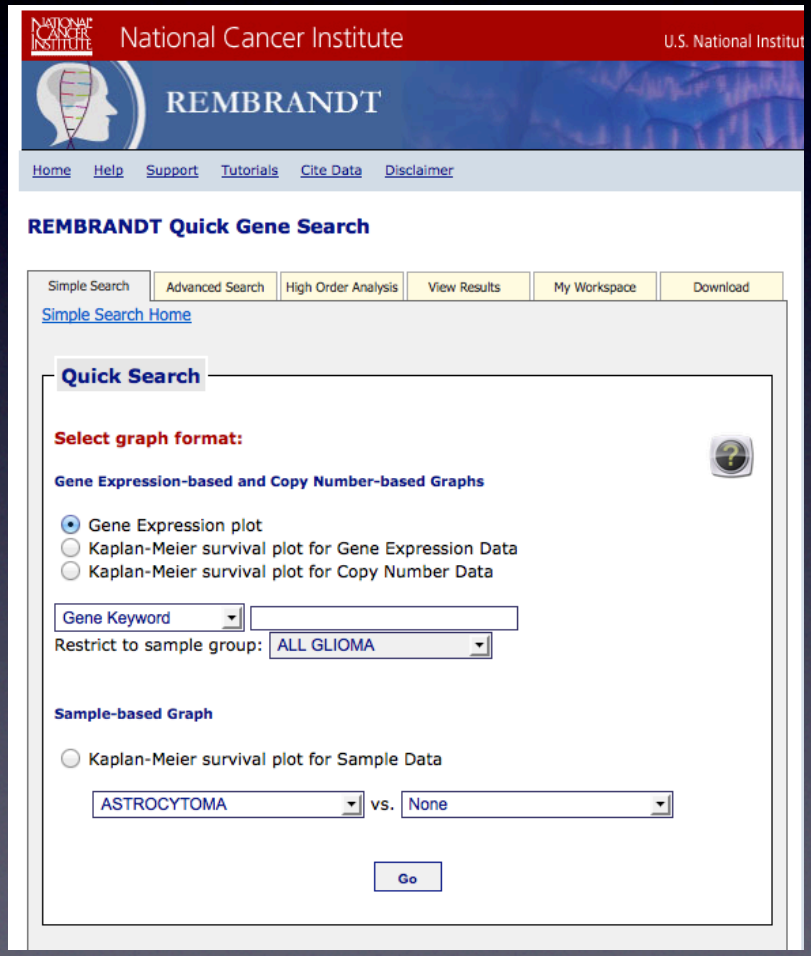
	T-test[1]	Linear regression[1]	Common
Affymetrix HT HG U133A	630	1749	**595**
Affymetrix Human Exon 1.0 ST	62	91	**40**
Agilent 244K G4502A	348	909	**313**
Common (U133A and G4502A)	**130**	**334**	**115**
Common (all three platforms)[2]	**17**	**40**	**14**

The last row shows the number of differentially expressed genes found in all three platforms.

[1]In each test, FDR≤0.05 threshold is applied.

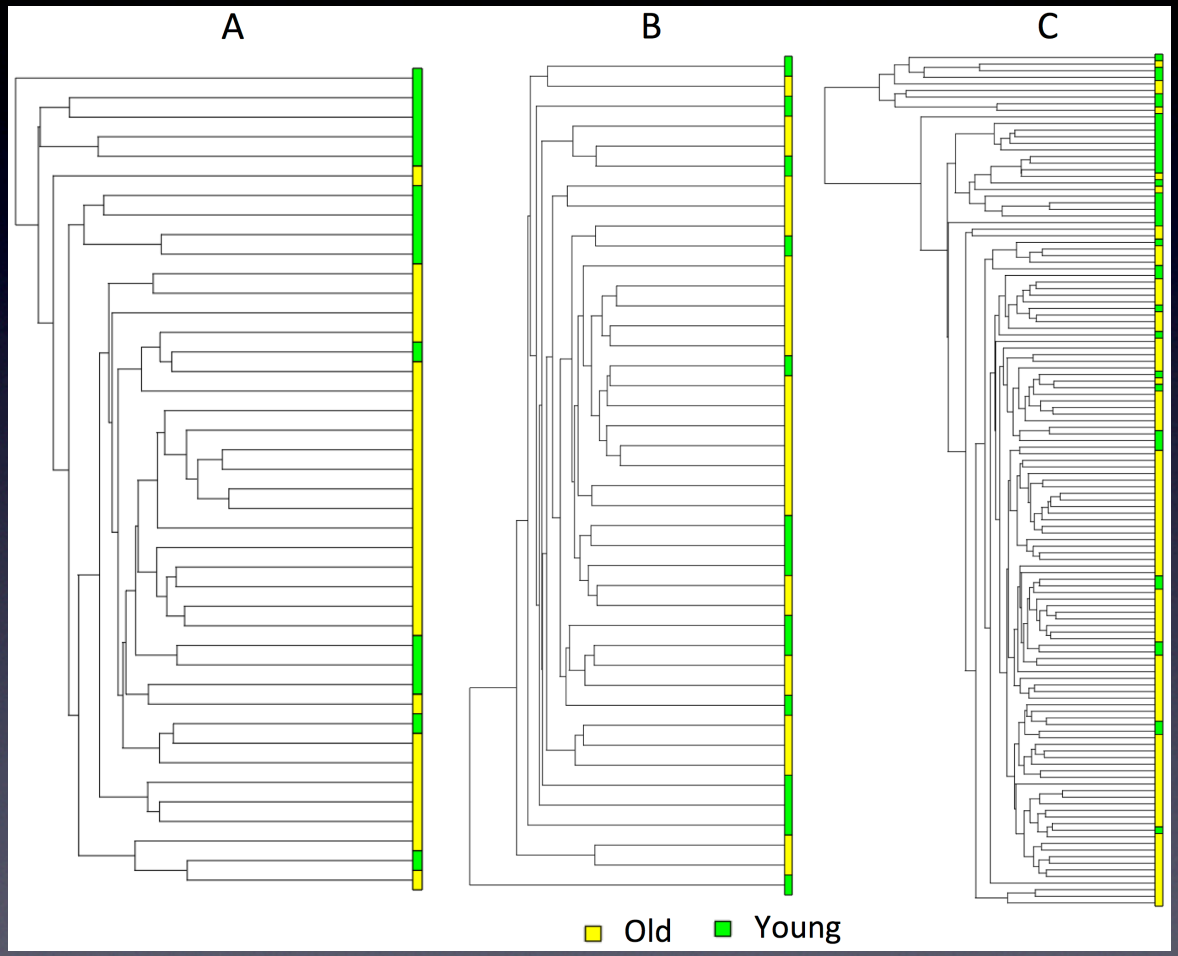
[2]Shows the number of differentially expressed genes found in all three platforms.

Validation of DEGs on external data (REMBRANDT)



https://caintegrator.nci.nih.gov/rembrandt/home.do

Validation of DEGs on external dataset



REMBRANDT
TCGA DEGs

REMBRANDT
All genes

TCGA
TCGA DEGs

Motif analysis

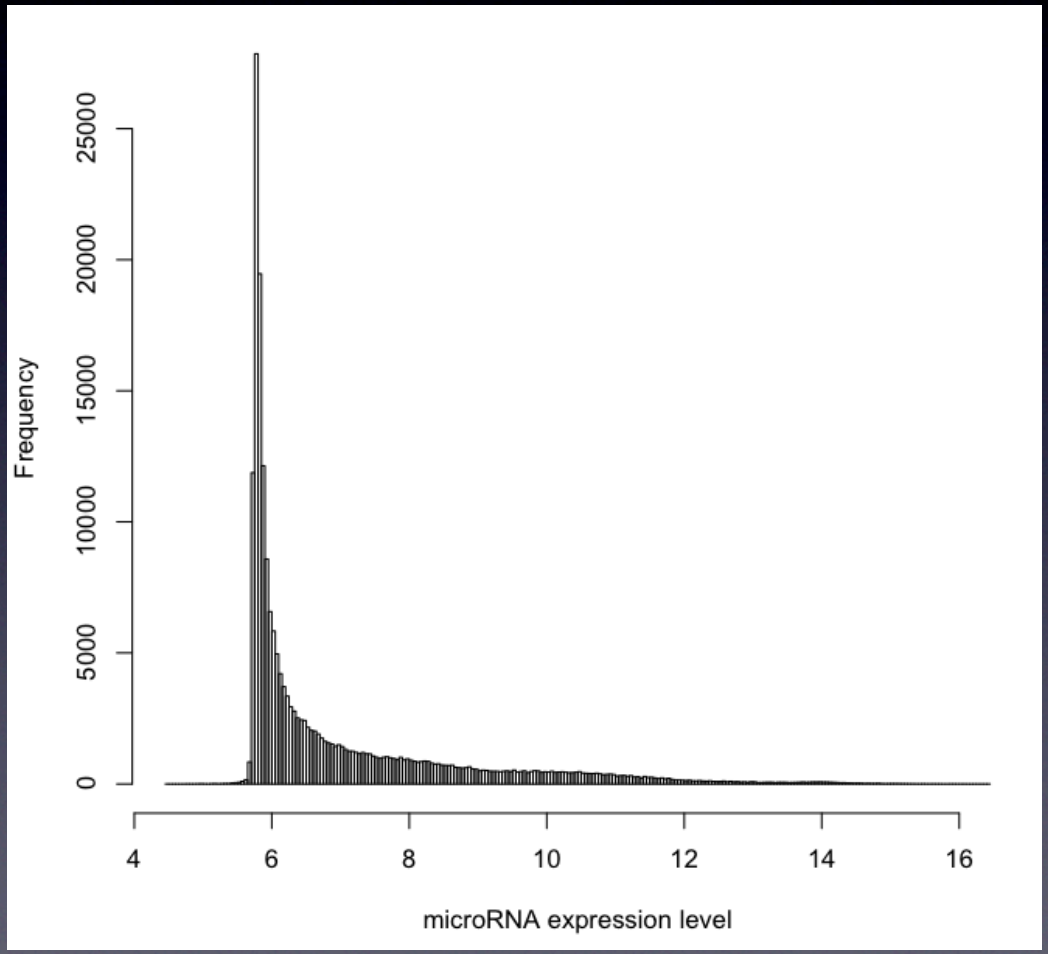
- Motif Enrichment analysis using JASPAR motifs and the PSCAN algorithm for promoter regions -450bp to + 50bp

TF Motif	FDR(BH)
Egr1	0.0025
INSM1	0.0030
MZF1_1-4	0.0157
PLAG1	0.0157
CTCF	0.0226
TFAP2A	0.0241
Mycn	0.0241
SP1	0.0241
Myc	0.0284
HIF1A::ARNT	0.0437

Functional analysis of DEGs

- DAVID results
- enrichment in several GO terms such as “response to hypoxia” (p-value,0.00123, enriched genes: VEGFA, SOD2, BNIP3, SLC11A2, EGLN3, PLOD2, NOL3, and ALDOC);
- “vasculo-genesis” (p-value,0.088, enriched genes: VEGFA, NTRK2, and QKI)
- VEGFA is a gene that has role in angiogenesis. It is up-regulated in old GBMs.

TCGA miRNA Histogram



Greg

Differentially expressed miRNAs

- Ranked-based linear regression
- FDR < 0.05

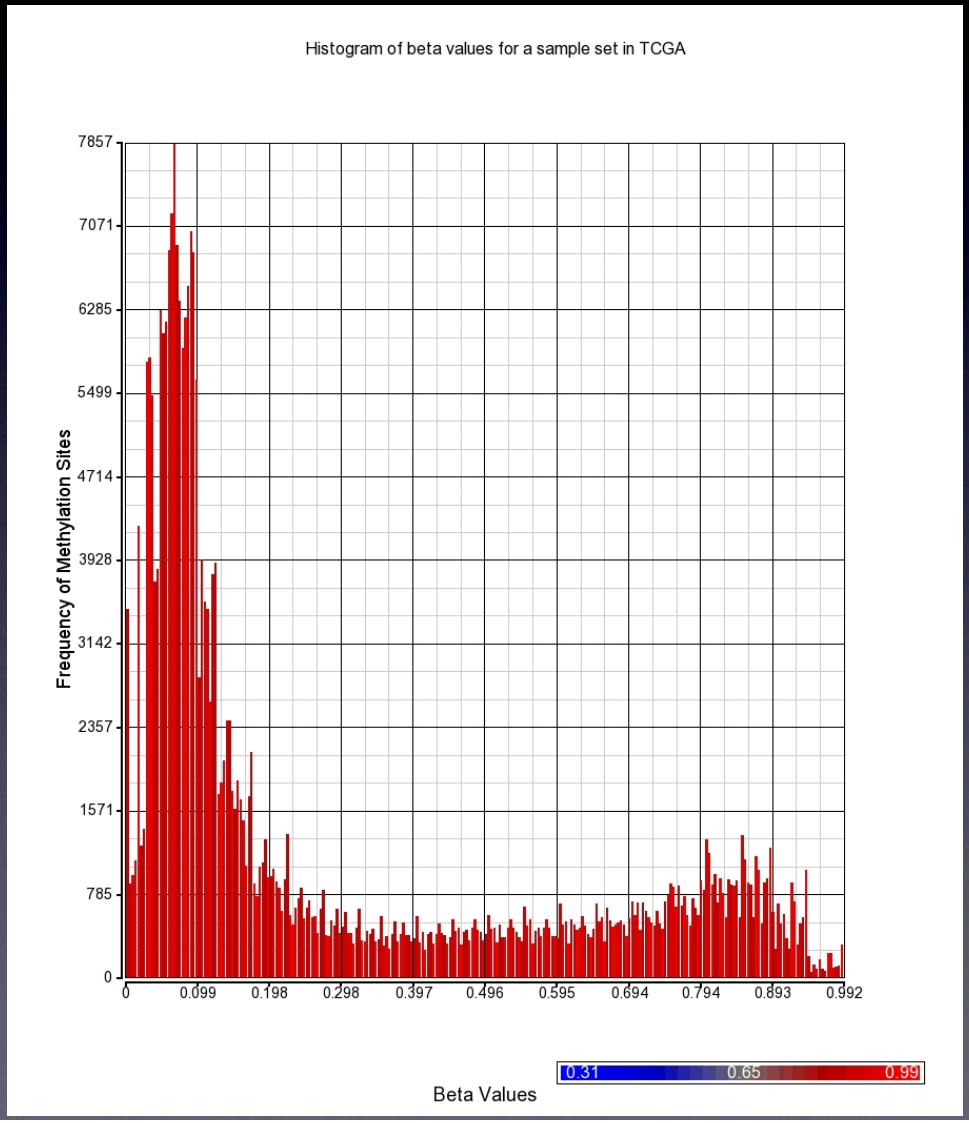
Differentially expressed microRNAs

- 19 miRs, all downregulated in old
- 172 experimentally validated target genes (mirWalk database)
- 7 of them are upregulated in old
  - LOX, VEGFA, DDIT4L, BCL6, MAF, NR2F1, SOX2

TCGA Methylation Data

- Infinium HumanMethylation27 Platform by Illimuna
- Each value in the data set : M/(U+M) where M and U are the signal intensities for methylated and unmethylated bead types, respectively.

Histogram of beta values



Differentially methylated genes (DMGs)

- Ranked-based linear regression is used to compute DMGs
- 389 age-specific DMGs were found
- 98% of them are hypermethylated in old

Human Molecular Genetics, 2011, Vol. 20, No. 6 **1164–1172**
doi:10.1093/hmg/ddq561
Advance Access published on January 7, 2011

Distinct DNA methylation changes highly correlated with chronological age in the human brain

Dena G. Hernandez[1,3,†], Michael A. Nalls[1,†], J. Raphael Gibbs[1,3], Sampath Arepalli[1], Marcel van der Brug[4], Sean Chong[1], Matthew Moore[1], Dan L. Longo[2], Mark R. Cookson[1], Bryan J. Traynor[1] and Andrew B. Singleton[1,*]

[1]Laboratory of Neurogenetics and [2]Lymphocyte Cell Biology Unit, National Institute on Aging, Baltimore, MD, USA [3]Department of Molecular Neuroscience and Reta Lila Weston Laboratories, Institute of Neurology, UCL, Queen Square House, London WC1N 3BG, UK and [4]Department of Molecular and Integrative Neurosciences, The Scripps Research Institute, Jupiter, FL, USA

Received July 15, 2010; Revised December 7, 2010; Accepted December 26, 2010

Methylation at CpG sites is a critical epigenetic modification in mammals. Altered DNA methylation has been suggested to be a central mechanism in development, some disease processes and cellular senescence. Quantifying the extent and identity of epigenetic changes in the aging process is therefore potentially important for understanding longevity and age-related diseases. In the current study, we have examined DNA methylation at >27 000 CpG sites throughout the human genome, in frontal cortex, temporal cortex, pons and cerebellum from 387 human donors between the ages of 1 and 102 years. We identify CpG loci that show a highly significant, consistent correlation between DNA methylation and chronological age. The majority of these loci are within CpG islands and there is a positive correlation between age and DNA methylation level. Lastly, we show that the CpG sites where the DNA methylation level is significantly associated with age are physically close to genes involved in DNA binding and regulation of transcription. This suggests that specific age-related DNA methylation changes may have quite a broad impact on gene expression in the human brain.

OPEN ACCESS Freely available online PLoS GENETICS

Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain

J. Raphael Gibbs[1,2☯], Marcel P. van der Brug[1,3☯], Dena G. Hernandez[1,2☯], Bryan J. Traynor[1], Michael A. Nalls[1], Shiao-Lin Lai[1,2], Sampath Arepalli[1], Allissa Dillman[1], Ian P. Rafferty[1], Juan Troncoso[4], Robert Johnson[5], H. Ronald Zielke[5], Luigi Ferrucci[6], Dan L. Longo[7], Mark R. Cookson[1]*, Andrew B. Singleton[1]*

1 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, **2** Department of Molecular Neuroscience and Reta Lila Weston Laboratories, Institute of Neurology, University College London, London, United Kingdom, **3** Department of Molecular and Integrative Neurosciences, The Scripps Research Institute, Jupiter, Florida, United States of America, **4** Division of Neuropathology, Department of Pathology, School of Medicine, Johns Hopkins University, Baltimore, Maryland, United States of America, **5** National Institute of Child Heath and Human Development Brain and Tissue Bank for Developmental Disorders, University of Maryland Medical School, Baltimore, Maryland, United States of America, **6** Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, United States of America, **7** Lymphocyte Cell Biology Unit, Laboratory of Immunology, National Institute on Aging, National Institutes of Health, Baltimore, Maryland, United States of America

Abstract

A fundamental challenge in the post-genome era is to understand and annotate the consequences of genetic variation, particularly within the context of human tissues. We present a set of integrated experiments that investigate the effects of common genetic variability on DNA methylation and mRNA expression in four human brain regions each from 150 individuals (600 samples total). We find an abundance of genetic *cis* regulation of mRNA expression and show for the first time abundant quantitative trait loci for DNA CpG methylation across the genome. We show peak enrichment for *cis* expression QTLs to be approximately 68,000 bp away from individual transcription start sites; however, the peak enrichment for *cis* CpG methylation QTLs is located much closer, only 45 bp from the CpG site in question. We observe that the largest magnitude quantitative trait loci occur across distinct brain tissues. Our analyses reveal that CpG methylation quantitative trait loci are more likely to occur for CpG sites outside of islands. Lastly, we show that while we can observe individual QTLs that appear to affect both the level of a transcript and a physically close CpG methylation site, these are quite rare. We believe these data, which we have made publicly available, will provide a critical step toward understanding the biological effects of genetic variation.

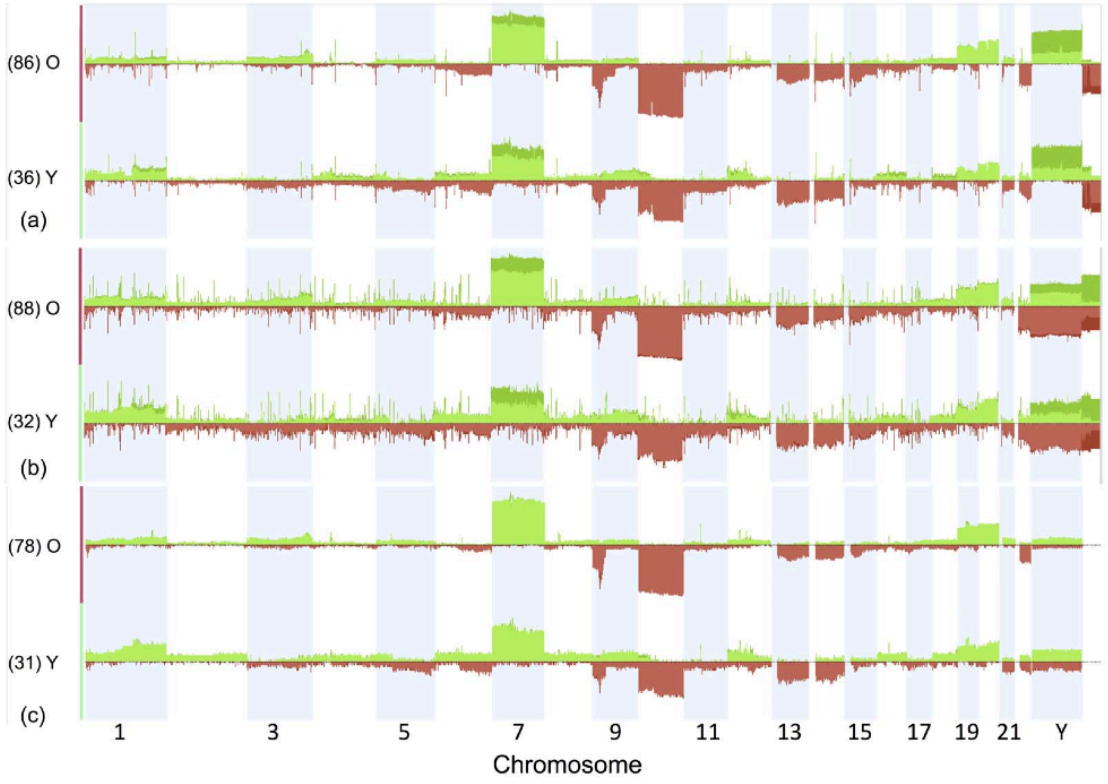
Singleton Data

- All methylation and clinical data are available on NCBI GEO (GSE15745)
- About 130 patients
- Samples from four regions of the brain
- Total 506 samples

DMGs

- Hypermethylated genes with aging in normal brain were filtered out
- 184 of them are uniquely hypermehylation in older GBMs
- 18 of them (Fisher's exact test p-value < 1.27e-05) are found to be associated in cancer through methylation

TCGA copy number alteration data



Differentially altered genes (DAGs)

- *Comparison* function (Nexus)
- q-value < 0.05

Age-specific CNA

	CNA (SNP6)
More frequently deleted in old	722
More frequently amplified in old	321
More frequently amplified in young	1
Total	1044

Eight hypermethylated genes in old (heterozygous deletion)

RASGEF1A C10orf47 HHEX PLCE1 FRMD4A SVIL ITGA8 PDLIM1

MOLECULAR AND CELLULAR BIOLOGY, May 2010, p. 2120–2134 Vol. 30, No. 9
0270-7306/10/$12.00 doi:10.1128/MCB.01511-09

PRH/Hhex Controls Cell Survival through Coordinate Transcriptional Regulation of Vascular Endothelial Growth Factor Signaling[▽]

Peter Noy,[1] Hannah Williams,[2] Anyaporn Sawasdichai,[2]
Kevin Gaston,[2] and Padma-Sheela Jayaraman[1]*

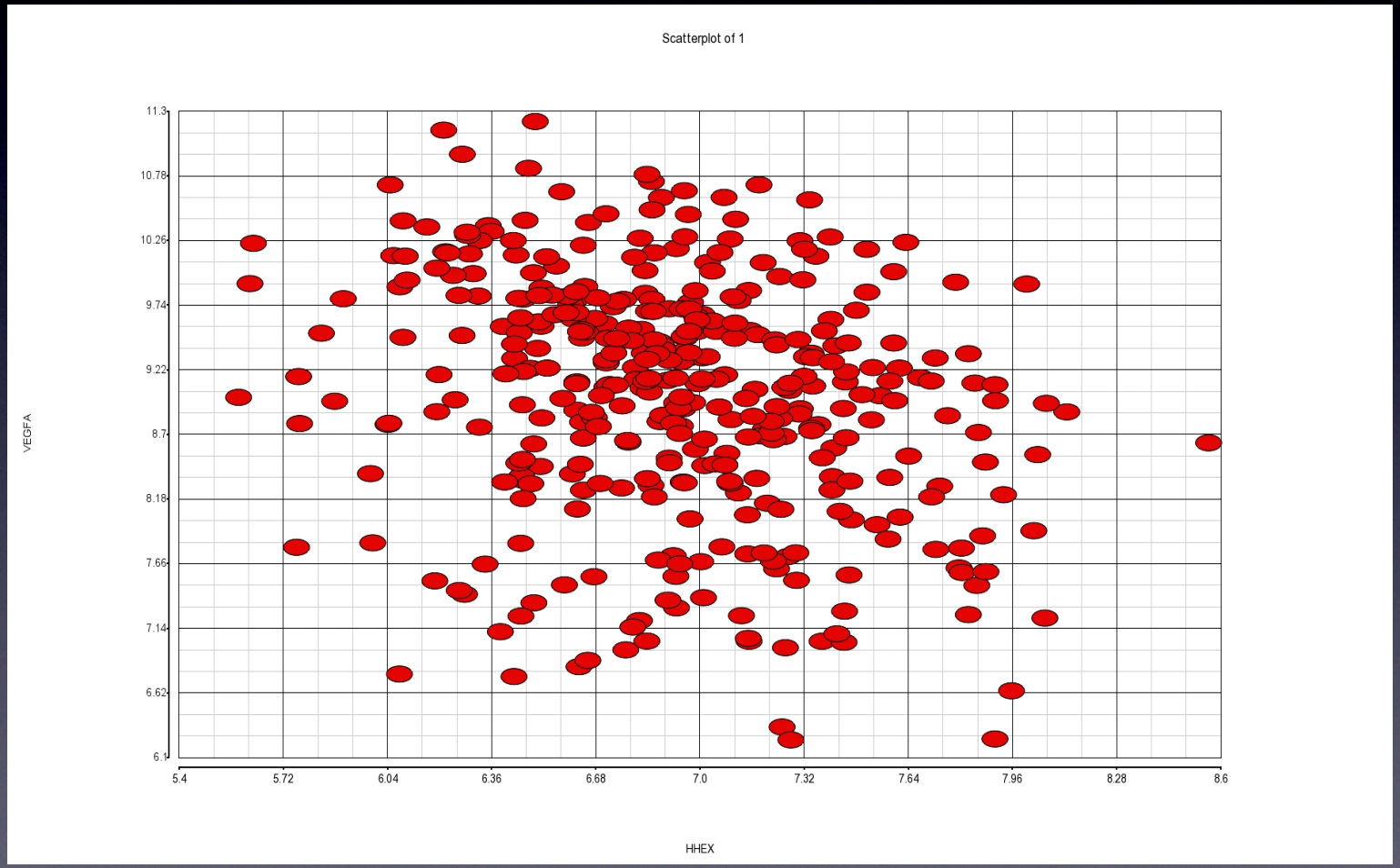
Institute for Biomedical Research, Birmingham University Medical School, Edgbaston, Birmingham B15 2TT, United Kingdom,[1] and Department of Biochemistry, University of Bristol, University Walk, Bristol BS8 1TD, United Kingdom[2]

Hypothesis

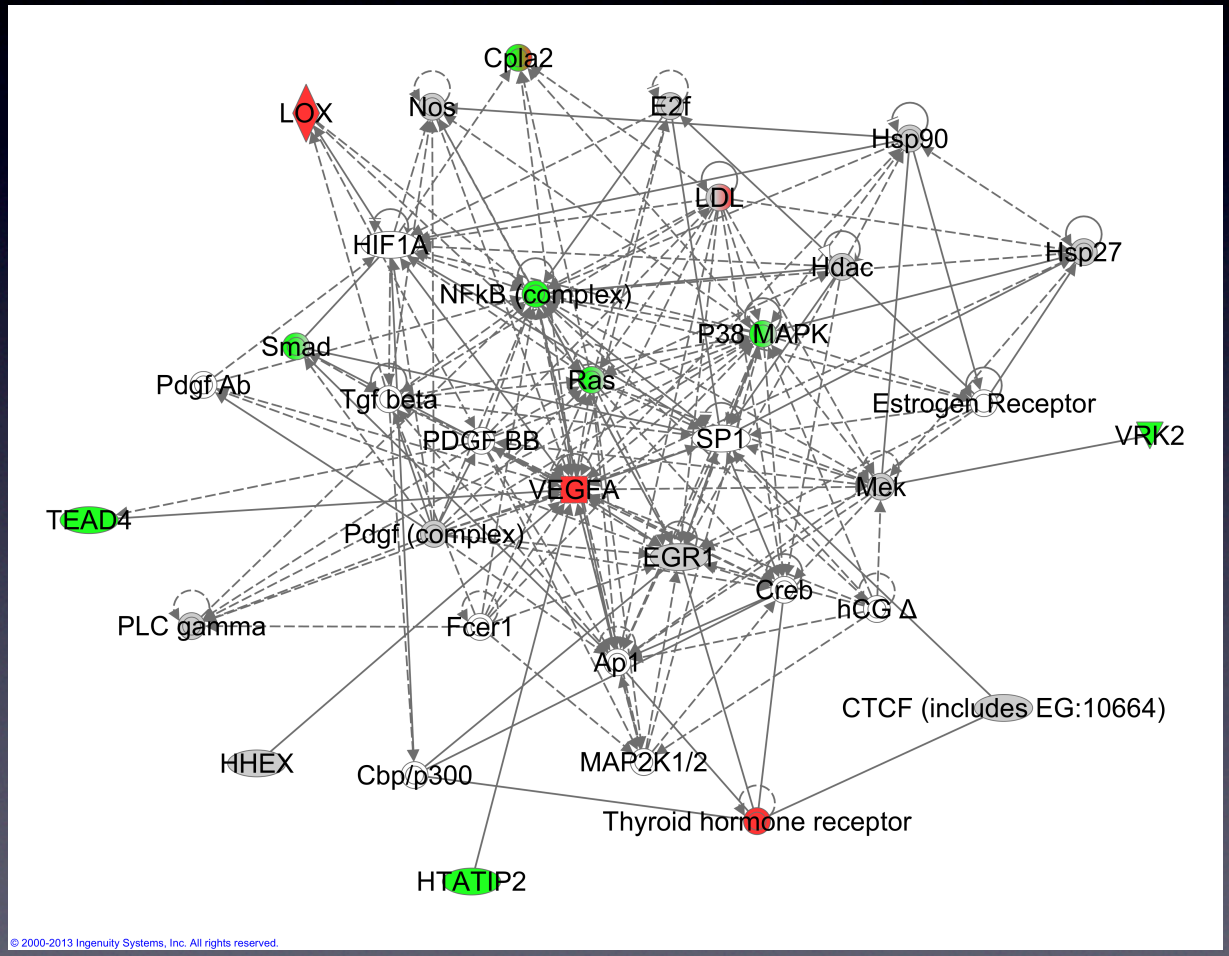
Vasculogenic/angiogenic upregulation in old GBMs compared to young GBMs through inhibition of VEGF by HHEX

HHEX vs. VEGF

Scatterplot (Exon data)



Network of angiogenesis



Summary

- Age is an independent significant prognostic factor of survival in GBMs
- Older GBMs have higher activity of angiogenesis
- Treatments that inhibit angiogenesis work better in old GBMs
- Activity of angiogenesis in older GBMs might be through deletion or hypermethylation of HHEX