**Identification of Common genes and proteins in Alzheimer's Disease, Multiple Sclerosis
and Duchenne Muscular Dystrophy using in-silico methods.**

Subhankar Kumar Singh

Souvik Chakraborty

Dr. Tarasankar Maiti

Anuska Chakraborty

Department of Physiology

Bhairab Ganguly College,

Department of Life Sciences

Presidency University.

ABSTRACT

Alzheimer's is a type of dementia symptom that slowly worsens over some time. In its early stages, memory loss is mild, but with late-stage Alzheimer's, the patient loses the ability to carry on a conversation and respond to their environment. Multiple sclerosis (MS) is an autoimmune disease of the central nervous system (CNS) characterized by chronic inflammation, demyelination, gliosis, and neuronal loss. Duchenne muscular dystrophy (DMD) is one of the most severe forms of inherited muscular dystrophies. It is the most common hereditary neuromuscular disease and does not exhibit a predilection for any race or ethnic group. In this study, several gene expression study data were analyzed and there were 557 Differentially Expressed Genes (DEGs) in all three chosen Datasets. A protein-protein interaction network was created using STRING and CytoHubba plug-in was used to identify the top ten genes which are POLR2A, SETD2, EFTUD2, RBM25, PRPF40A, CDK13, BPTF, THOC2, SNRNP70, and SCAF11. Online software Enrichr was used for Gene Ontology and KEGG pathway enrichment analysis to find out the biological process, molecular function, cellular component, and the pathways that are commonly affected in these diseases.

INTRODUCTION

Alzheimer's disease (AD) is the most common kind of dementia, accounting for more than two-thirds of dementia cases in adults aged 65 and up. Alzheimer's disease is a progressive neurological illness that affects behavioural and cognitive processes such as memory, understanding, language, attention, thinking, and judgement. In the United States, it is the sixth biggest cause of death. Early onset (beginning before 65 years of age) is uncommon, occurring in less than 10% of Alzheimer's disease patients. Alzheimer's disease has no cure, although there are therapies that can help with some of the symptoms.ⁱ

MS (Multiple Sclerosis), the most common neurological disease, is an autoimmune-mediated condition that affects the central nervous system (CNS) and frequently causes severe physical or cognitive incapacitation as well as neurological issues in young adults. The principal causes of myelin sheath degradation include multifocal zones of inflammation caused by focal T-lymphocytic and macrophage infiltrations, as well as oligodendrocyte mortality. Inflammatory cells and their products, demyelinated and transected axons, and astrogliosis in both white and grey matter result in the formation of CNS plaques. These lesions can interfere with nerve impulse transmission and cause problems.ⁱⁱ

Due to abnormalities in the dystrophin gene, Duchenne muscular dystrophy (DMD) is an X-linked hereditary neuromuscular illness.ⁱⁱⁱ It's characterised by progressive muscle weakening and wasting caused by a lack of dystrophin protein, which causes skeletal and cardiac muscle degeneration. The most common neuromuscular condition, Duchenne muscular dystrophy (DMD, MIM 310200), affects up to 1/3600 male births globally. It's caused by mutations in the dystrophin gene on the X chromosome, and the symptoms don't appear until later in life. When the initial symptoms develop, the typical age of diagnosis is four years.^{iv}

In Alzheimer's individuals, the genetic biomarkers are **AβPP**, **PSEN1**, **PSEN2**^v. Whereas in Multiple Sclerosis and Duchenne Muscular Dystrophy the genetic biomarkers are **CD163**, **YKL40**, and **FCER2**, **MMP9**^{vi}.

In this study, datasets for Alzheimer's disease, Multiple sclerosis, and Duchenne Muscular Dystrophy were obtained from the Gene expression omnibus database, and have been analyzed to find out the common genes between these diseases and what are the pathways which have been affected.

METHODS

Dataset Selection

Gene expression Omnibus (GEO) by the National Center of Biotechnology Information (NCBI) is an internationally acclaimed online database for high-throughput sequencing data, microarray, hybridization array data.^{vii}

Geo was searched for datasets as "Alzheimer", "Multiple Sclerosis", "Duchenne Muscular Dystrophy", "expression profiling by array", "Homo sapiens". For Alzheimer we got the dataset with its GEO accession number GSE1297, then for Multiple Sclerosis, we got the dataset with GEO accession number GSE21942 and our last dataset for this study is of Duchenne Muscular Dystrophy with GEO accession number GSE38417. All the datasets which are used in this study are available online and no actual human experiments were performed by the author. In this study total of 41 patients, samples, and 29 normal control subjects were present. All three studies used microarrays for data collection. Microarray is a chip-based technology in which 1000s of nucleic acids are bound to a surface and are used to measure the relative concentration of nucleic acid sequences in a mixture.^{viii}

Table I - Dataset information for Microarray Dataset obtained from GEO. There are in total 13 Alzheimer patients and 8 Normal samples as the control in dataset GSE1297, 12 Multiple Sclerosis patients, and 15 Normal samples as the control in dataset GSE21942, 16 Duchenne Muscular Dystrophy patients, and 6 Normal Sample as the control in dataset GSE38417.

Dataset	Number of Patient	Number of Control
GSE1297	13	8
GSE21942	12	15
GSE38417	16	6
Total	41	29

Differential Gene Expression Analysis

The over 3 datasets were investigated utilizing the GEO2R tool (accessible at <http://www.ncbi.nlm.nih.gov/geo/geo2r/>) which is given by the Gene Expression Omnibus. GEO2R is online software that processes gene expression values and outputs a table of two differentially expressed genes (DEGs) between two user-defined groups i.e., patients and control in 3 datasets. After running GEO2R, p values were found out using the 't' test. Fold changes of each gene were calculated with the help of the GEO2R tool. Fold change is represented by logFC values which tell us about the magnitude of gene expression change. Suppose, as a result, a value of 3.2 is 2 to the power 3.2 ($2^{3.2}$). This suggests that gene expression levels for this gene are 5 times higher in patients than in normal subjects. The GEO2R tool was also used to verify the normal distribution of samples. There were no outliers in these datasets.

Fig. I

Fig. II

Fig. III

Fig. I- Gene expression value distribution for Dataset GSE1297, Fig. II- Gene expression value distribution for dataset GSE21942, Fig. III- Gene expression value distribution for Dataset GSE38417.

GEO2R tool was used for verifying a normal distribution of samples. No outliers were present in these datasets. After obtaining the list of DEGs, they were processed in Excel. DEGs with a p-value greater than 0.05 were removed and the rest of the DEGs were sorted into two categories, which are overexpressed and those which are under-expressed.

Fig. IV- Venn diagram representing the DEGs of all three datasets i.e., GSE1297, GSE21942, GSE38417

Identification of common genes

A software named Fun Rich was used for creating Venn Diagram through which we can get the common genes between 3 datasets.^{ix}

Network Analysis

STRING (<https://string-db.org/>) was used to examine the DEGs in the final list. It's utilised to build protein networks that communicate with one another. STRING is a gene ontology process and pathway search engine that also calculates statistically significant Gene Ontology processes and pathways. STRING provided a tab-separated value file, which was then imported into Cytoscape. Cytoscape is a free programme that combines biomolecular interaction networks with high-throughput expression data. The gene hub was discovered using a network analyzer, a Cytoscape plugin called CytoHubba.

Enrichment Analysis of DEGs

Gene enrichment analysis was performed to find out DEGs using the online software called Enrichr^x (accessible at <https://maayanlab.cloud/Enrichr/>) for Gene Ontology (GO) function and also for KEGG pathway enrichment.

RESULTS

Three gene expression profiles (GSE1297, GSE21942 and GSE38417) were analyzed with the help of the GEO2R tool and the list of DEGs was downloaded. p-value was set at 0.05 and a log FC value of 0.5 was set as the cut-off criteria. There were 897 genes in the dataset GSE1297, 493 genes in the dataset GSE21942 and 2998 genes in dataset GSE38417. Software Fun Rich is used to perform Venn analysis and then DEGs were identified. 557 genes were found to be significantly differentially expressed among all three mentioned datasets. In these datasets, 23 genes are differentially expressed in all datasets and the name of those genes are SNORD10, SNORA48, SNORA67, EIF4A1, SCAF11, ELOVL5, ZBTB20, COX20, PRRC2C, PRPF40A, ZFP36L1, PTPRK, SWAP70, NKTR, RPL37A, ACP5, WIPF1, SCAF4, NOTCH2NL, RUFY3, DOCK10, YME1L1, PRMT2, NSUN6, AHSA2, CDK13. Now the 23 common genes were imported to the STRING tool. The disconnected genes were concealed by clicking 'hide disconnected nodes in the network' inside the 'network analysis' option and the medium confidence option was selected in the minimum required interaction score option. Total 503 nodes and 1626 edges are present in the network (Fig. V). For finding the hub genes out of the 23 genes, the whole network was imported to the Cytoscape offline tool. Inside the CytoHubba plug-in, the top 10 hub genes were found by selecting confidence (score) cut-off at 0.40 and maximum additional interaction at 0.

Table II – This table includes top 10 Hub Genes with their full name and Degree. CTNNB1 gene has the maximum interaction with other Hub genes.

GENE	FULL NAME	DEGREE
CTNNB1	Catenin Beta 1	76.0
PTPRC	Protein Tyrosine Phosphate Receptor Type C	54.0
KRAS	Kirsten Rat Sarcoma Virus	49.0
CD44	Cyclin Dependant Kinase 44	45.0
RPS27A	Human Ubiquitin Fusion Gene UBA80	38.0
PIK3R1	Phosphoinositide-3-Kinase Regulatory Subunit 1	37.0
HNRNPA1	Heterogenous Nuclear Ribonuclear Protein 1	32.0
PLEK	Pleckstrin	32.0
CD86	Cyclin Dependant Kinase 86	32.0
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	31.0

The GO and KEGG analysis was carried out using the online tool Enrichr. After submitting the hub genes inside the ontology option in the Enrichr, Enrichr provided several options like GO-biological process (GO-BP), GO-molecular function (GO-MF), GO-cellular component (GO-CC). GO analysis showed that the DEGs were enriched in GO-BP such as negative regulation of programmed cell death, negative regulation of apoptotic process, positive regulation of interleukin-2 production, regulation of immunoglobulin production, hematopoietic progenitor cell differentiation, positive regulation of telomere maintenance via telomerase, positive regulation of immunoglobulin production, positive regulation of telomere maintenance via telomere lengthening, positive regulation of production of molecular mediator of immune response, cytokine-mediated signaling pathway (Fig. VIA). GO-MF analysis showed that DEGs were especially phosphatase binding, protein phosphatase binding, ErbB-3 class receptor binding, neurotrophin TRK receptor binding, neurotrophin TRKA receptor binding, G-rich strand telomeric DNA binding, histone methyltransferase binding, hyaluronic acid binding, transmembrane receptor protein tyrosine kinase adaptor activity, single-stranded telomeric DNA binding, (Fig. VIB). For GO-CC analysis, the DEGs are enriched with focal adhesion, cell-substrate junction, bounding membrane of organelle, cytoplasmic vesicle membrane, mitochondrial outer membrane, organelle outer membrane, basolateral plasma membrane, phosphatidylinositol 3-

kinase complex class I, bleb, beta-catenin-TCF complex. The above-discussed hub genes are upregulated in Alzheimer, Multiple Sclerosis, Duchenne Muscular Dystrophy.

Fig. V - STRING protein-protein interaction network. In this network, there are 503 nodes and 1626 edges. Lines represent interaction and the circles represent common genes. The color of the lines represents the type of interaction.

A.

negative regulation of programmed cell death (GO:0043069)
 negative regulation of apoptotic process (GO:0043066)
 positive regulation of interleukin-2 production (GO:0032743)
 regulation of immunoglobulin production (GO:0002637)
 hematopoietic progenitor cell differentiation (GO:0002244)
 positive regulation of telomere maintenance via telomerase (GO:0032212)
 positive regulation of immunoglobulin production (GO:0002639)
 positive regulation of telomere maintenance via telomere lengthening (GO:1904358)
 positive regulation of production of molecular mediator of immune response (GO:0002702)
 cytokine-mediated signaling pathway (GO:0019221)

B.

negative regulation of programmed cell death (GO:0043069)
 negative regulation of apoptotic process (GO:0043066)
 positive regulation of interleukin-2 production (GO:0032743)
 regulation of immunoglobulin production (GO:0002637)
 hematopoietic progenitor cell differentiation (GO:0002244)
 positive regulation of telomere maintenance via telomerase (GO:0032212)
 positive regulation of immunoglobulin production (GO:0002639)
 positive regulation of telomere maintenance via telomere lengthening (GO:1904358)
 positive regulation of production of molecular mediator of immune response (GO:0002702)
 cytokine-mediated signaling pathway (GO:0019221)

C.

focal adhesion (GO:0005925)
 cell-substrate junction (GO:0030055)
 bounding membrane of organelle (GO:0098588)
 cytoplasmic vesicle membrane (GO:0030659)
 mitochondrial outer membrane (GO:0005741)
 organelle outer membrane (GO:0031968)
 basolateral plasma membrane (GO:0016323)
 phosphatidylinositol 3-kinase complex, class I (GO:0097651)
 bleb (GO:0032059)
 beta-catenin-TCF complex (GO:1990907)

D.

Fig. VI- GO and KEGG pathway enrichment analysis using Enrichr. A- Top 10 enriched biological processes in hub genes. The x-axis represents the number of genes and the y axis represents the biological process. B- Top 10 enriched molecular functions in hub genes. The x-axis represents the number of genes, and the y axis represents the molecular function. C- Top 10 enriched cellular components in hub genes. The x-axis represents the number of genes and the y axis represents cellular components. D – Top 10 enriched KEGG pathways for hub genes. The x-axis represents the number of genes, and the y axis represents the KEGG pathway names.

Fig. VII- Network of top ten hub genes using Cytoscape software. The color represents the degree of connectivity, the red color represents the highest degree, the orange color represents intermediate degree, and the yellow represents the lowest degree.

DISCUSSION

Through this study, some key genes have been identified, which were previously not reported. Genes such as CTNBN1, PTPRC, KRAS, CD44, RPS27A, PIK3R1, HNRNPA1, PLEK, CD86, SMARCA44 were previously not known to regulate Alzheimer, Multiple Sclerosis and Duchenne Muscular Dystrophy in common.

Through previous studies, we know that beta-amyloid and tau levels in cerebrospinal fluid (CSF) acts as biomarkers of Alzheimer's^{xi}. HLA-DRB1 acts as a biomarker of Multiple Sclerosis shown in a study^{xii}. whereas there is no evidence of a common gene between Alzheimer, Multiple Sclerosis and Duchenne Muscular Dystrophy. Several pathways, biological functions associated with the above-mentioned upregulated genes were also enriched in the KEGG and GO enrichment analysis. So this article can help further for common genetic biomarkers of these diseases.

The study's limitations include the lack of data normalization and the lack of differentiation between male and female patients, as well as normal controls, where there can be considerable differences in differentially expressed genes concerning gender. For these limitations, further study is required to find out more potential common genetic biomarkers for these three diseases.

REFERENCES

- Al-Khalili Szigyarto, Cristina, and Pietro Spitali. "Biomarkers of Duchenne Muscular Dystrophy: Current Findings." *Degenerative Neurological and Neuromuscular Disease* Volume 8 (January 2018): 1–13. <https://doi.org/10.2147/DNND.S121099>.
- "Alzheimer's Disease." *Dialogues in Clinical Neuroscience* 2, no. 2 (June 2000): 91–100. <https://doi.org/10.31887/DCNS.2000.2.2/asschachter>.
- Axisa, Pierre-Paul, and David A. Hafler. "Multiple Sclerosis: Genetics, Biomarkers, Treatments." *Current Opinion in Neurology* 29, no. 3 (June 2016): 345–53. <https://doi.org/10.1097/WCO.0000000000000319>.
- Barrett, Tanya, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, et al. "NCBI GEO: Archive for Functional Genomics Data Sets—Update." *Nucleic Acids Research* 41, no. D1 (November 26, 2012): D991–95. <https://doi.org/10.1093/nar/gks1193>.
- Bumgarner, Roger. "Overview of DNA Microarrays: Types, Applications, and Their Future." In *Current Protocols in Molecular Biology*, edited by Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, and Kevin Struhl, mb2201s101. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013. <https://doi.org/10.1002/0471142727.mb2201s101>.
- Chung, Jeffrey, Andrea L. Smith, Sarah C. Hughes, Gabriela Niizawa, Hoda Z. Abdel-Hamid, Edwin W. Naylor, Timothy Hughes, and Paula R. Clemens. "Twenty-Year Follow-up of Newborn Screening for Patients with Muscular Dystrophy: NBS in DMD and BMD." *Muscle & Nerve* 53, no. 4 (April 2016): 570–78. <https://doi.org/10.1002/mus.24880>.
- Compston, Alastair, and Alasdair Coles. "Multiple Sclerosis." *The Lancet* 372, no. 9648 (October 2008): 1502–17. [https://doi.org/10.1016/S0140-6736\(08\)61620-7](https://doi.org/10.1016/S0140-6736(08)61620-7).
- Di Gregorio, Eleonora, Barbara Borroni, Elisa Giorgio, Daniela Lacerenza, Marta Ferrero, Nicola Lo Buono, Neftj Ragusa, et al. "ELOVL5 Mutations Cause Spinocerebellar Ataxia 38." *American Journal of Human Genetics* 95, no. 2 (August 7, 2014): 209–17. <https://doi.org/10.1016/j.ajhg.2014.07.001>.
- Duan, Dongsheng, Nathalie Goemans, Shin'ichi Takeda, Eugenio Mercuri, and Annemieke Aartsma-Rus. "Duchenne Muscular Dystrophy." *Nature Reviews Disease Primers* 7, no. 1 (December 2021): 13. <https://doi.org/10.1038/s41572-021-00248-3>.
- Fagerberg, Linn, Björn M. Hallström, Per Oksvold, Caroline Kampf, Dijana Djureinovic, Jacob Odeberg, Masato Habuka, et al. "Analysis of the Human Tissue-Specific Expression by Genome-Wide Integration of Transcriptomics and Antibody-Based Proteomics." *Molecular & Cellular Proteomics: MCP* 13, no. 2 (February 2014): 397–406. <https://doi.org/10.1074/mcp.M113.035600>.

- . “Analysis of the Human Tissue-Specific Expression by Genome-Wide Integration of Transcriptomics and Antibody-Based Proteomics.” *Molecular & Cellular Proteomics: MCP* 13, no. 2 (February 2014): 397–406. <https://doi.org/10.1074/mcp.M113.035600>.
- Falzarano, Maria, Chiara Scotton, Chiara Passarelli, and Alessandra Ferlini. “Duchenne Muscular Dystrophy: From Diagnosis to Therapy.” *Molecules* 20, no. 10 (October 7, 2015): 18168–84. <https://doi.org/10.3390/molecules201018168>.
- Ghasemi, Nazem, Shahnaz Razavi, and Elham Nikzad. “Multiple Sclerosis: Pathogenesis, Symptoms, Diagnoses and Cell-Based Therapy.” *Cell J (Yakhteh)* 19, no. 1 (January 2017). <https://doi.org/10.22074/cellj.2016.4867>.
- Hoffman, Eric P., Robert H. Brown, and Louis M. Kunkel. “Dystrophin: The Protein Product of the Duchenne Muscular Dystrophy Locus.” *Cell* 51, no. 6 (December 1987): 919–28. [https://doi.org/10.1016/0092-8674\(87\)90579-4](https://doi.org/10.1016/0092-8674(87)90579-4).
- Housley, William J., David Pitt, and David A. Hafler. “Biomarkers in Multiple Sclerosis.” *Clinical Immunology* 161, no. 1 (November 2015): 51–58. <https://doi.org/10.1016/j.clim.2015.06.015>.
- Huynh, Rose Ann, and Chandra Mohan. “Alzheimer’s Disease: Biomarkers in the Genome, Blood, and Cerebrospinal Fluid.” *Frontiers in Neurology* 8 (March 20, 2017). <https://doi.org/10.3389/fneur.2017.00102>.
- Jones, Kelly A., Yue Luo, Lynn Dukes-Rimsky, Deepak P. Srivastava, Richa Koul-Tewari, Theron A. Russell, Lauren P. Shapiro, Anand K. Srivastava, and Peter Penzes. “Neurodevelopmental Disorder-Associated ZBTB20 Gene Variants Affect Dendritic and Synaptic Structure.” Edited by Xiangming Zha. *PLOS ONE* 13, no. 10 (October 3, 2018): e0203760. <https://doi.org/10.1371/journal.pone.0203760>.
- Kim, N. S., T. Kato, N. Abe, and S. Kato. “Nucleotide Sequence of Human CDNA Encoding Eukaryotic Initiation Factor 4A1.” *Nucleic Acids Research* 21, no. 8 (April 25, 1993): 2012. <https://doi.org/10.1093/nar/21.8.2012>.
- Kumar, Anil, Jaskirat Sidhu, Amandeep Goyal, and Jack W. Tsao. “Alzheimer Disease.” In *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2021. <http://www.ncbi.nlm.nih.gov/books/NBK499922/>.
- Ohno, Yusuke, Shota Suto, Masao Yamanaka, Yukiko Mizutani, Susumu Mitsutake, Yasuyuki Igarashi, Takayuki Sassa, and Akio Kihara. “ELOVL1 Production of C24 Acyl-CoAs Is Linked to C24 Sphingolipid Synthesis.” *Proceedings of the National Academy of Sciences of the United States of America* 107, no. 43 (October 26, 2010): 18439–44. <https://doi.org/10.1073/pnas.1005572107>.
- Pathan, Mohashin, Shivakumar Keerthikumar, Ching-Seng Ang, Lahiru Gangoda, Camelia Y.J. Quek, Nicholas A. Williamson, Dmitri Mouradov, et al. “FunRich: An Open Access Standalone Functional Enrichment and Interaction Network Analysis Tool.” *PROTEOMICS* 15, no. 15 (August 2015): 2597–2601. <https://doi.org/10.1002/pmic.201400515>.
- Shannon, P. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.” *Genome Research* 13, no. 11 (November 1, 2003): 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- Sherif, Fayroz F., Nourhan Zayed, and Mahmoud Fakhr. “Discovering Alzheimer Genetic Biomarkers Using Bayesian Networks.” *Advances in Bioinformatics* 2015 (August 23, 2015): 1–8. <https://doi.org/10.1155/2015/639367>.
- Suárez-Fariñas, Mayte, Michelle A. Lowes, Lisa C. Zaba, and James G. Krueger. “Evaluation of the Psoriasis Transcriptome across Different Studies by Gene Set Enrichment Analysis (GSEA).” Edited by H. Peter Soyer. *PLoS ONE* 5, no. 4 (April 20, 2010): e10247. <https://doi.org/10.1371/journal.pone.0010247>.
- Szklarczyk, Damian, Annika L. Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, et al. “STRING V11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets.” *Nucleic Acids Research* 47, no. D1 (January 8, 2019): D607–13. <https://doi.org/10.1093/nar/gky1131>.

GSE1597

Q502 P42

GSE12417

OSCI279

OSCI47

OSCI442

negative regulation of transcription, involved in cell cycle

negative regulation of apoptosis, positive APC activity

positive regulation of intracellular protein production (20-30%)

regulation of transcription, in protein case (2-3%)

transcriptional repression, cell cycle

positive regulation of telomere

positive regulation of cell

positive regulation of

positive regulation

regulation of

phosphatase binding (GO:0011381)

phosphatase binding (GO:0011381)

phosphatase binding (GO:0011381)

phosphatase binding (GO:0011381)

phosphatase binding (GO:0011381)

phosphatase binding (GO:0011381)

phosphatase binding (GO:0011381)

phosphatase binding (GO:0011381)

phosphatase binding (GO:0011381)

phosphatase binding (GO:0011381)

total reference data coverage

all data from January 1999 to 2000

coverage of individual data points

coverage of individual data points

coverage of individual data points

coverage of individual data points

coverage of individual data points

coverage of individual data points

coverage of individual data points

coverage of individual data points