Journal of Alzheimer's Disease 101 (2024) 813–822
DOI 10.3233/JAD-240174
IOS Press

Molecular Gene Expression Testing to Identify Alzheimer's Disease with High Accuracy from Fingerstick Blood

Bruce Seligmann[a,*], Salvatore Camiolo[a], Monica Hernandez[a], Joanne M. Yeakley[a], Gregory Sahagian[b] and Joel McComb[a]
[a]*BioSpyder Technologies, Inc., Carlsbad, CA, USA*
[b]*Neurology Center of Southern California, Carlsbad, CA, USA*

Handling Associate Editor: Junichi Iga

Accepted 12 July 2024
Pre-press 11 September 2024

Abstract.
Background: There is no molecular test for Alzheimer's disease (AD) using self-collected samples, nor is there a definitive molecular test for AD. We demonstrate an accurate and potentially definitive TempO-Seq® gene expression test for AD using fingerstick blood spotted and dried on filter paper, a sample that can be collected in any doctor's office or can be self-collected.
Objective: Demonstrate the feasibility of developing an accurate test for the classification of persons with AD from a minimally invasive sample of fingerstick blood spotted on filter paper which can be obtained in any doctor's office or self-collected to address health disparities.
Methods: Fingerstick blood samples from patients clinically diagnosed with AD, Parkinson's disease (PD), or asymptomatic controls were spotted onto filter paper in the doctor's office, dried, and shipped to BioSpyder for testing. Three independent patient cohorts were used for training/retraining and testing/retesting AD and PD classification algorithms.
Results: After initially identifying a 770 gene classification signature, a minimum set of 68 genes was identified providing classification test areas under the ROC curve of 0.9 for classifying patients as having AD, and 0.94 for classifying patients as having PD.
Conclusions: These data demonstrate the potential to develop a screening and/or definitive, minimally invasive, molecular diagnostic test for AD and PD using dried fingerstick blood spot samples that are collected in a doctor's office or clinic, or self-collected, and thus, can address health disparities. Whether the test can classify patients with AD earlier then possible with cognitive testing remains to be determined.

Keywords: Alzheimer's disease, dementia, diagnostic classifier, fingerstick blood, gene expression signature, health disparity, Parkinson's disease, self-test, TempO-Seq, whole blood

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia, affecting more than 55 million people worldwide with more than 10 million new cases/year.[1] It is estimated that three-fourths of those with AD worldwide have not been diagnosed, and on average it can take over 2 years for a diagnosis of AD when one is made.[2] 6.5 million Americans live with AD, with a risk of developing AD of 1 in 5 for women, and 1 in 10 for men.[3] The prevalence of AD is higher among non-Hispanic Black and Hispanic Americans, populations which experience health dis-

*Correspondence to: Bruce Seligmann, 5928 Pascal Ct., Suite 100/150, Carlsbad, CA 92008, USA. Tel.: +1 520 237 3077; E-mail: bseligmann@biospyder.com.

parities reducing access to healthcare. This may result in under-diagnosis, particularly at earlier stages of disease.

Estimates of AD prevalence are based on cognitive testing because there is no FDA approved definitive molecular test. By "definitive" we mean a test, which if positive for AD, does not require further testing other than to stratify patients and assess their progression. Currently, a definitive diagnosis of AD requires a battery of tests including clinical phenotype of AD and biomarker evidence of AD pathology. Thus, when a patient presents with clinical manifestations of possible AD, a definitive diagnosis is based on assessment of patient history, neurological exam, objective cognitive/functional assessment to establish a clinical phenotype commonly associated with AD, scans (MRI, CT, PET) to rule out hemorrhages, strokes, etc., a positive immunoassay for tau and amyloid-β (Aβ), and may include positive FDG PET/CT and Aβ PET scans.[4] This series of tests requires access to a medical specialist and numerous laboratory visits. While immunoassays measuring Aβ and tau proteins from cerebrospinal fluid or plasma have received FDA clearance as *in vitro* diagnostic devices, they are not sufficient alone to diagnose AD.[5–10] Cognitively unimpaired subjects can have a positive tau or amyloid immunoassay test and be classified as at-risk for progression to a dementia but may never develop clinical manifestations of AD within their lifetime.[4] Thus, the intended use of these tests are for patients already being evaluated for AD.[6]

Not only is there no definitive molecular test for AD, but there is no molecular test that uses a minimally invasive sample permitting collection in any doctor's office or self-collection that can be used to screen persons to identify those who should seek further diagnosis for AD. Furthermore, the number of AD patients greatly outweighs the number of neurologists, making it important to have a test that can be caried out on a sample that can be collected in any doctor's office or clinic, to identify those who should be seen for further diagnosis.[11,12] Finally, the inaccessibility of many patients to any doctor, much less a neurologist, leads to health disparities, creating a need for a classification test that can use a self-collected sample without having to see any doctor first. We report a classifier test for AD that uses a sample that can be collected in any doctor's office or clinic, or which can be self-collected to inform subjects that they should see a neurologist for diagnosis. While there is the potential for this test to provide a definitive diagnosis, it has clinical utility even if it is not definitive because of how easy it is to collect the sample, and it has utility as a research use assay generally in the field of dementia research.

There are many reports of gene expression signatures for AD and PD that have utilized RNA extracted from whole blood or from white cells or were derived bioinformatically from databases of such samples.[13–28] All these used blood obtained by venipuncture. While most did not correlate their results to biomarker assays, one did, demonstrating a good correlation (classifying 24 of 28 positive patients) to the CSF amyloid biomarker test.[18] Another reported an accuracy of 74–77% predicting AD in patients with mild cognitive impairment (MCI) two years prior to diagnosis of AD.[21] Thus, there is good evidence that a whole blood test can be used to classify patients with AD.

TempO-Seq® is an assay platform that uses crude sample lysates to measure the expression of focused sets of genes up to the whole transcriptome.[29] Seeking to develop a minimally invasive test, we pursued the use of dried blood spots on filter paper prepared from a fingerstick as the assay sample. Others have profiled RNA extracted from blood spotted on filter paper, but not from AD patients.[30] Because the TempO-Seq assay does not require RNA extraction, we chose to directly test punches of dried blood collected on filter paper. The TempO-Seq assay has also been shown to provide the same quality data from highly degraded RNA (RIN 3.0) as from high quality RNA samples, a feature we believed would be beneficial profiling RNA within the spotted blood samples.[29] This TempO-Seq Dried Blood Spot (TempO-SeqDBS) assay was used to determine whether patients clinically diagnosed with AD could be identified and differentiated from asymptomatic controls and patients clinically diagnosed with PD.

MATERIALS AND METHODS

We tested cohorts of patient samples to identify a gene signature and algorithm that classified patients with AD and differentiated them from controls and PD. While subtypes of AD based on gene expression profiling have been described, our objective was to identify a test that could classify patients as having AD regardless of disease subtype.[31] Thus, we used an approach that identified an AD classifier signature that was in common to all patients in the training set. We subsequently retrained and retested the algo-

rithm on independent cohorts of samples to confirm the ability of the test to classify patients.

Patients being seen at the Neurology Center of Southern California with a clinical diagnosis of AD or PD and asymptomatic controls, were consented under an IRB-approved protocol (Solutions IRB, protocol ID 0417). The clinical diagnosis of AD followed the clinical guidance published by the National Institute on Aging and Alzheimer's Association (NIA-AA), published in 2011, and involved a thorough medical evaluation, including a review of the patient's medical history, physical examination, cognitive testing using, e.g., the Mini-Mental State Examination or Montreal Cognitive Assessment, neurological evaluation including MRI/CT imaging to exclude other causes of the cognitive functional loss, and lab testing to exclude metabolic diseases that could account for the cognitive function loss. The clinical diagnosis of PD followed the guidance published by the UK Parkinson's Disease Society Brain Bank for PD, and was based on medical history, review of symptoms, neurological and physical exam, identifying the presence of characteristic motor symptoms of bradykinesia, rigidity, postural instability, and resting tremor. Blood was obtained by a fingerstick, 2-3 drops spotted on Whatman filter paper to generate a spot with a diameter of ∼1.5 cm, air dried, de-identified, and then shipped to BioSpyder and tested. Two areas, each 1.6 mm diameter, were collected using a hand punch and placed into a microplate well, with four replicate wells per patient sample (providing 4 within patient biological replicates). Samples were lysed by heating at 95°C for 10 min in 2 μL of Denaturation Buffer covered with 10 μL of mineral oil. The assay was initiated by adding 2 μL of an Annealing Mix containing detector oligos from the commercial TempO-Seq Whole Transcriptome Whole Blood assay v2.1 (measuring all ∼21,000 genes) and incubated at 70°C for 10 min, followed by a temperature ramp to 45°C over 50 min, then held at 45°C for 16 h before being cooled to 25°C. 20 μL of Whole Blood Nuclease Mix was added to the aqueous layer (achieved by adding 20 μl nuclease buffer and then transferring the supernatant into a fresh microplate well and adding concentrated nuclease mix), and samples incubated for 1.5 h at 37°C before adding 24 μL of Ligation Mix to each well, incubating 1 h at 37°C, then incubating for 15 min at 80°C, before dropping the temperature to 25°C. 10 μL of each well containing the sample-specific ligated products, were transferred for amplification into TempO-Seq PCR Primer Plates containing sample-specific indexed universal forward and reverse PCR primers and PCR mix. After 30 cycles, the ligated detector oligos for each sample were uniquely indexed and sequencing primers incorporated into the PCR adduct, to prepare a sequencible product. 5 μL of the PCR product from each sample well were pooled together into a sequencing library, purified using the Macherey-Nagel NucleoSpin Gel and PCR Cleanup Kit (catalog number 740609.50) and quantified by absorbance using the 260/230 and 260/280 ratios and Qubit fluorescence. The filter paper TempO-Seq processing kits with the reagents described above are now commercially available, for use with the content of any TempO-Seq assay.

Sequencing was carried out on an Illumina NextSeq to an average depth of ∼6 M reads/sample. Demultiplexing was carried out on the sequencer to correctly associate individual FASTQ files (and ultimately counts) with each indexed sample. Commercial TempO-SeqR™ software was then used for alignment to generate count tables and evaluate the sequencing run. Quality control metrics for analysis were repeatability between replicates, determined using a Pearson correlation, with any samples among the four replicates that had an $R < 0.9$ correlation to the remaining replicates being removed from analysis as an outlier, while requiring at least 3 remaining replicates/sample. We calculated the NSig80 (the number of probes that capture 80% of the signal) and the NCov10 (the number of probes receiving less than 10 reads) for each sample, and performed analysis to assure that distributions of these metrics were consistent between the classes of controls, AD, and PD.

RESULTS

Components of whole blood have previously been shown to interfere with certain steps in the TempO-Seq protocol, an interference we were successful in overcoming to establish the Tempo-SeqDBS protocol.[32] We discovered that after spotting on filter paper and drying, interferences were eliminated by the TempO-SeqDBS protocol, permitting gene expression to be profiled. The reproducibility of whole transcriptome data measured by the TempO-SeqDBS assay from different replicate regions within the spot tested for each donor is shown as Pearson correlations, calculated using expression of all genes as a variable so that significance values could be calculated for the overall comparison rather than at a

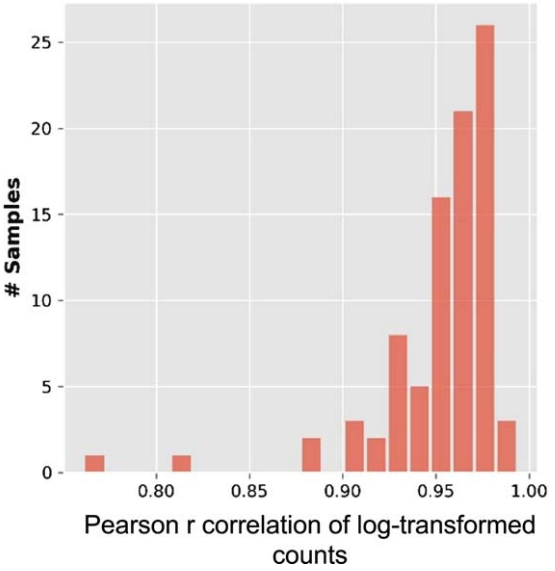

Fig. 1. Histogram showing the distribution of the Pearson r correlation between replicates as a measure of reproducibility. Four different areas of the filter paper spotted with finger stick blood from each donor were tested as replicates. The Pearson correlations between all the possible pairwise comparisons between the replicates for each donor sample were calculated after log transforming the data, and the results depicted as a histogram of the number of samples grouped into each Pearson coefficient value.

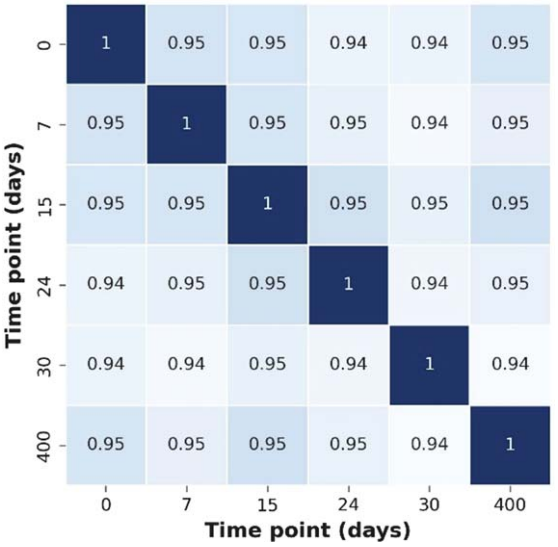

Fig. 2. Pearson r correlation of log-transformed counts to quantify reproducibility of gene expression measurement from the same donor, different days. Samples from the same donor were spotted and tested on different days as controls over a period of 400 days. The replicates for the same time point were averaged and the average log transformed values were correlated using the Pearson correlation, comparing days 0, 7, 15, 24, 30, and 400. The Pearson correlations are indicated as the values of the heatmap.

gene level, between all possible pairwise comparisons between the replicates for each donor sample after log transforming the data (Fig. 1). The coefficients reported in Fig. 1 were all significant with a $p < 0.05$. Assay repeatability between finger stick samples collected and tested from the same subject on different days over a period of 400 days is depicted in Fig. 2. In this case, all the replicates for the same time point were averaged, log transformed, values were correlated, and the Pearson correlation for each comparison plotted.

The TempO-SeqDBS whole transcriptome assay was used to test AD, PD, and control samples to identify signatures of differentially expressed genes to be subsequently used to identify classification algorithms for AD and PD. A first cohort (cohort A) of samples was profiled using the whole transcriptome assay. The majority, 24 controls, 28 AD, and 27 PD passed the quality control replicate sample metric (see Methods). A set of genes was identified that provided a differentiating signature by performing pairwise comparison between the three classes of patients. To account for the presence of multiple dementia subtypes in our dataset, we contrasted different "classes" (AD to control, PD to control, and AD to PD) of samples by testing hundreds of random subsets and performing differential gene expression analyses (DESeq2). The genes that proved to be differentially expressed in at least 33% of the tests ($n = 770$ differentially expressed genes for control vs AD, control versus PD, AD versus PD, Table 1) were used as a signature differentiating the three classes of patients.

Before proceeding to identify classification algorithms, the identified signature genes were grouped into pathways to explore/establish the translational significance of the differentially expressed signature genes. We performed an over-representation analysis in well-annotated biological processes (p value threshold of 0.05). The predominant pathways were immune response pathways (Fig. 3), with the three most significantly differentially expressed pathways related to the involvement of neutrophils in the immune and inflammatory response.

We used different machine learning approaches, together with the 770 gene signature, to build algorithms that differentiated control, AD and PD. Methods used were k-nearest neighbors, Random forest, Support vector machine (SVM) with either linear, polynomial or Radial Basis Function (rbf) kernel and Extreme Gradient Boosting (XGboost)

Table 1
List of the genes within the 770 and 68 gene signatures

770 Gene signature													
FCRL2	H1F0	SERPINA1	FOLR3	NPIPA1	CSF2RB	JCHAIN	BLVRB	MYZAP	ADTRP	RUNX3	CD6	ZDHHC2	CSF3R
JUNB	GNAI2	BTG2	OGFRL1	CYP4F3	FCGR2A	RPL4	SH2D1B	SLC6A8	FCMR	FCGR3B	GYPC	RGCC	CTDSP1
CRISPLD2	S100A12	DUSP1	MPIG6B	FMN1	MMP25	KCNJ15	BCL2A1	ODC1	RHOG	IER2	IGHA2	ABCA7	DCAF12
EIF1AY	HLA-DRA	CXCR1	IGLV3-21	AMD1	RAB11FIP1	TUBB1	FYN	RPL13	HEBP1	CCL4	RAB31	RPL19	DDX5
BSG	OST4	RPL3	HAGH	CD8A	HLA-DRB1	HIST1H4K	MAP4K1	ARHGAP15	TREM1	FYB1	PDZK1IP1	UBB	DEFA3
UTY	SEPTIN7	RBM38	WNK1	HLA-DRB4	ALDOA	CD3D	ADM	NFAM1	HNRNPC	PROK2	FGL2	KCNE3	FCGR3A
TRBC1	FAM126B	FOS	CD79A	SRSF5	USP48	CD22	SELENBP1	CXCR2	IDS	BST2	GIMAP5	S100A6	FCMR
HLA-DRB5	TNFSF13B	TCL1A	OAS3	OGT	MXD1	VPS8	SHISA4	PVALB	NBPF9	CCNI	TAP2	HBA2	FPR1
VMP1	GNA13	IFI30	GZMB	ATP6V0C	XPO6	ANXA2	ZNF417	AIF1	PLCB2	MGAT4A	RPS4X	OR1S2	FTL
NAGK	SLC25A39	SIRPB1	SERPING1	GNLY	VCAN	CST3	FCN1	SIRPA	TPCN2	KLF3	FXYD5	ARHGEF40	GABARAPL2
HAL	TMCO1	CEP85L	CA1	TRDC	RPL12	LCP1	SNAP23	HLA-C	TMCC3	TIMP2	C6orf62	PAIP2	GADD45B
NBPF12	PLXNB2	RPL23	RGS18	BCL2	YWHAZ	UBN1	SMCHD1	RNF182	PGGHG	FAM153A	SLC25A37	RPS2	GNAS
KLF4	FPR1	ALAS2	PARP1	CXCL8	TNFAIP8	TAF1D	AZU1	OSTF1	CD163	PRDX6	RNF213	VCL	GNLY
DENND4A	ZBTB1	GCA	PFN1	MAPK1	LRP1	IGHV3-23	NQO2	CASP1	TRGC1	UCP2	PSMF1	TENM1	GPX1
CAVIN2	MUC20	VSIR	IFIT1	MKRN1	TSPAN5	IL1R2	LMBRD1	UBA52	RPL36	NKG7	SMIM5	TCP11L2	GYPC
APEX1	ANXA11	UBALD2	USP12	VIM	B2M	ALDH2	CTSD	ITGB3	RAC2	CORO1B	CD33	AP1S2	HBA1
MMP9	DCUN1D1	SLC35A2	GABARAPL2	CD3E	IFI6	NEAT1	DGKA	NPRL3	MSN	CNOT1	INKA2	WASHC4	HBG1
RALGPS2	AC011462.1	WLS	S100B	USP4	NLRC5	FOSB	CTSB	ALOX5AP	RNF130	TSC22D3	SPTB	KRT5	HDGF
STRADB	NBEAL2	RPS23	NCF2	DDX3Y	FTH1	GBP5	FCGR3A	NKX3-1	NUP50	RPL28	TMEM154	BCL2L1	HLA-A
SORL1	SOCS3	HNRNPH1	HCK	IFIT2	MPEG1	RASSF2	RNASET2	IER5	TOMM7	ZNF728	IL6R	AGO4	HLA-C
SOX6	YPEL3	ISG15	AC124319.1	CD53	PABPC1	RPL9	MT2A	HLA-DQB1	MRC2	BNIP3L	LRRK2	SLC6A6	IFIT1
CCL3L1	N4BP2L2	HIST1H4E	TMSB10	SLC11A1	PHC2	GPX1	RPL18	CD3G	EVI2A	BCL6	PHOSPHO1	ZDHHC18	IGF2R
AHSP	TRIM25	CD68	ITGAM	EMC3	KDM5C	PECAM1	VPS13B	SH3BGRL	ADGRG3	PTGS2	DCAF12	CEACAM4	IGLV3-21
HK3	EPB42	MAP2K3	STAT1	UBE2L6	OCLN	NACA2	XK	ORAI2	ADGRE2	AGTRAP	BZW1	ABCB10	ITGB2
EMP3	RPL35A	ELF1	CD58	PLEKHG2	EPB41	MTRNR2L9	HLA-DRB3	WASHC1	S100A8	C9orf78	CD74	TYMP	LST1
HIST1H2BO	DICER1	CDC34	CARD8	TMEM176B	TM9SF2	CDC42	LGALS3	STEAP4	LITAF	KDM4C	STMP1	DEFA1	LYZ
MAX	CD226	TUBA1A	ATP6V0E1	LPIN2	TAGLN2	MICAL2	FCGR1B	HBZ	RPL23A	IGHG1	SH3BGRL3	LEF1	MAP2K3
SLFN5	RPS27	IGF2R	ZFAND6	TRIM22	PPP3R1	GATA1	GIMAP7	GMPR	SLC16A3	FBXO7	PCED1A	CTDSP1	MGAM
S100P	FBXL5	CA4	MXI1	CTSS	AGAP6	S100A4	DDT	VTI1B	MTRNR2L8	SEC62	UBXN6	SDCBP	MMP25
LTA4H	LILRA3	PIM2	MSR1	RPS6	FTL	APOBEC3A	MYL9	MED18	APLP2	NRGN	TMCC2	SRPRA	MNDA
VNN2	CRIP1	IRF1	SERF2	RPL21	KLF1	SLC25A28	RNF149	P2RX5-TAX1BP3	GLRX5	ABI3	HDGF	LAPTM5	MTRNR2L9
CD247	CR1	RGS2	NPM1	THNSL2	MX1	ADGRE5	PINK1	IFI44	ABCC4	LFNG	KAT2B	FAM210B	MYO1F
DGLUCY	LAP3	PPM1F	RPL26	HNRNPD	IL10RA	SLC38A5	AATK	TBC1D10C	HSP90AB1	KRT14	ASAH1	MEFV	NCF2
LYL1	PTGS1	GZMH	PYCARD	SRP14	LILRA4	CDR1	AQP9	M6PR	SIGLEC10	HIST1H1C	S100A9	LYZ	NEAT1
EIF4G2	TCEA1	MYL12B	CISD2	ZFC3H1	DENND2D	STX7	YBX1	RPL22	ANXA1	IGLC2	PGM5	HIST1H1E	NKG7
GADD45B	C15orf54	CMTM6	RPIA	SCPEP1	CYP27A1	IFIT1B	NBPF14	FCER1G	ZNF141	LAMP2	JUN	TMSB4X	NPRL3
MME	CSF1R	SRRM2	SPOPL	TYROBP	SOD2	GNAQ	PSMB9	ESPN	HVCN1	CMTM1	TRANK1	NCOA4	PDZK1IP1
HBA1	GNAS	SLC35C2	JAK1	SMG1	NIBAN3	IGHD	HSPB1	RESF1	ISCA1	TRGC2	LY6E	PIM1	PGGHG
EIF2S3B	IFITM1	PI3	TAPBP	RAB18	DGAT2	SLC7A5	RIOK3	FAM153B	EIF3L	OSBP2	P2RY13	HLA-A	PIP4K2A
ATP5MG	ARF6	FPR2	RGS10	BTNL3	MT1L	RAB2B	MGME1	HSPA6	RAI2	STAT6	POTEF	CDC42SE1	PRF1
TSC22D4	JUND	UBE2O	CFL1	NBPF8	HSPA8	CALM2	SLC44A2	GIMAP4	RPS8	IQGAP1	EFHD2	IFITM3	PROK2
NBPF26	SELL	ARHGAP9	DDX39B	LCK	RNF10	ADSS	ACOX1	TFDP1	ARPC3	YY1AP1	OTUB1	PPDPF	PSMF1
PPM1A	SRGN	CD46	CMPK2	HNRNPA1	ORMDL3	GNS	PLSCR1	MDGA1	PPP1CB	PTBP1	GANC	MDM4	RFK
ITGAX	VAMP3	ATF6B	OGA	SLC38A1	IGSF6	COTL1	LILRB3	IVNS1ABP	SRSF7	NOSIP	CD83	CD52	RNASET2
PGD	RGPD1	ACTB	ITGB1	FCAR	NMI	IFITM2	CRTAP	BCL11B	TXNIP	KIF20B	GUK1	TMEM123	S100A6
ATG16L2	SPOCK2	TPST1	UHMK1	MALAT1	ACTN1	MANBA	SULF2	F13A1	ZYX	PILRA	GSPT1	PRR13	S100A9
MX2	PITHD1	CD37	SRGAP2	MYO1F	PPBP	CD300E	PNISR	FBXW7	NTSR1	GID4	RPS16	GUCD1	S100P
VNN3	GZMA	NAGA	POTEJ	CD5	GPR146	TRIM58	KLRC2	IFIT5	SP110	CCR2	DEFA1B	PSAP	SEC62
ABHD18	TPT1	TMEM30A	NPIPB11	DDX5	PIP4K2A	HBG1	ACAP2	IGHM	CPSF1	HIST1H2BD	ANK1	ARL6IP1	SLC2A3
EVI2B	PADI4	TMOD1	EIF2AK1	FAM104A	RPS27A	CHCHD2	ZNF83	SIRPB2	HIPK3	ICAM3	APAF1	IGF2BP2	SORL1
AOAH	MTRNR2L6	WSB1	MARCH8	SPARC	CD36	MBNL1	SMARCA2	FAM117A	FOXO3	RHOH	SIGLEC9	**68 Gene**	TENT5C
NT5C3A	FBXL13	CFD	OPTN	ARHGDIB	LILRA5	ACSL1	TMEM164	S100A11	TUBB2A	BTF3	R3HDM4	**Signature**	TOMM7
FCER2	NINJ2	RFK	ITGB2	TLN1	TANK	PRF1	SRSF2	RPS4Y1	DMTN	AKR1E2	RBM39	ACTB	TRBC1
RPS3	RSRP1	TCF7	FECH	RHOQ	SEC14L1	INSIG1	CD8B	FCGR2C	RNF11	MPP1	RBBP4	ALAS2	TRIM22
HLA-E	PTBP3	UBE2H	UIMC1	FGFR1OP2	NCF4	ITGA2B	SF3B1	DPM2	ITGAL	TRIM10	RBM33	ARHGAP9	TRIM58
C4orf3	PRPF4B	TIMM10	TRGV4	SMIM1	SAT1	SLC15A4	CCL4L2	MGAM	HEMGN	CYBB	VENTX	ARHGDIB	UBB
ELOB	LST1	FLNA	MNDA	KDM5D	RAF1	EGR1	PAK2	SH3KBP1	FCGR1A	RPS20	NUSAP1	B2M	VSIR
ADIPOR1	SLC2A3	GBP2	LDHB	FOXO4	DEFA3	MYL4	HMOX1	STK17B	NPIPB4	ACTG1	LGALS2	BNIP3L	YBX1
TMEM50A	CSF3R	TNFRSF10C	PRKAR1A	GTF2IP1	RPS15A	HBM	HCAR3	ARRDC3	FKBP8	MS4A1	NBPF19	C9orf78	YBX3
ZEB2	OAZ1	THEMIS2	CCR7	PLEK2	TCIRG1	YBX3	CHI3L1	RSAD2	PXN	UBE2W	TENT5C	CRIP1	YWHAZ

and for all of these models hyper-parameters were tuned using a Random Grid approach. Randomly sampled subsets accounting for 80% of the initial dataset were used to train these classifiers, which were then tested on the remaining 20% of the samples. For each classifier an area under receiver operating characteristic (ROC) curve (AUC-ROC) was calculated for classification of both AD and PD. The Support Vector Machine (SVM) model (kernel = linear, gamma = 0.1, c = 0.001) resulted in the best performance for the classification of both AD (ROC-$AUC_{AD} = 0.81$), and for PD (ROC-$AUC_{PD} = 0.88$), as shown in Fig. 4. This model was used without further retraining to classify an independent cohort of 22 AD patients (cohort B), 18 of which were predicted as AD while the remaining 4 were predicted as PD. The two cohorts were pooled and the same SVM model was retrained using 80% of the samples and tested with the remaining 20%. The AUC for AD increased to 0.86 (Fig. 5), and the AUC for PD remained 0.87. The model misclassified one control classified as AD, one AD as control, and one PD as AD thus providing an accuracy of 97% across all 106 samples. Thus, while the average age of controls was 67, of AD was 78, and PD was 68, and there were a percentage of patients with co-morbidities, these

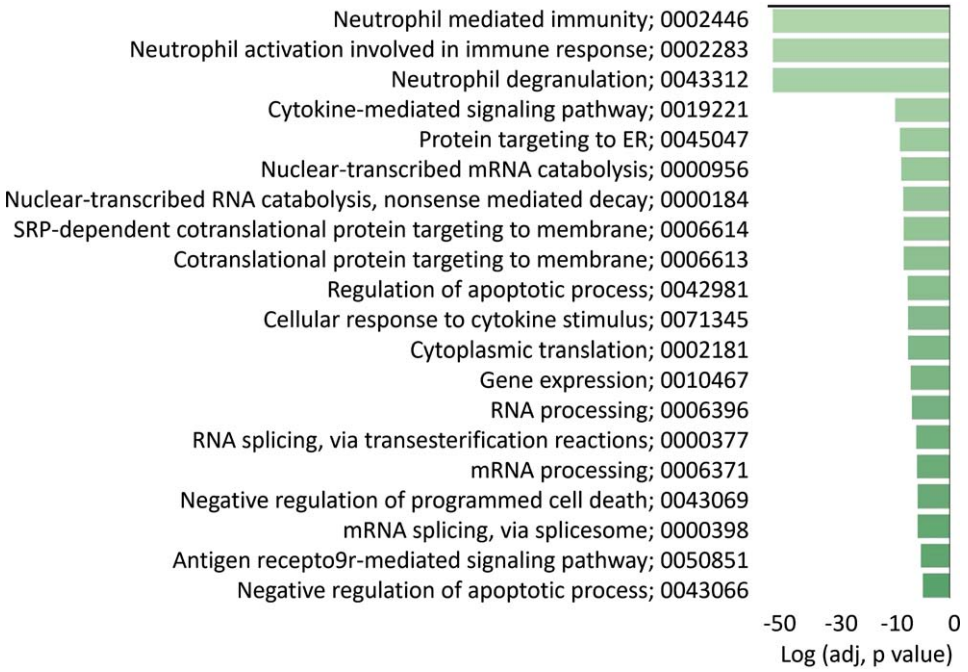

Fig. 3. Gene expression pathway analysis. The differentially expressed genes in the AD classifier signature were used to carry out analysis of what molecular pathways were differentially expressed. Gene set enrichment was performed using the Gene Ontology database for biological processes with a p value threshold of 0.05. The 20 most significant identified pathways are depicted, ranked by log adjusted p value for each.

factors did not appear to have significantly impacted classification and the 97% accuracy of the test.

The 770 gene signature was further investigated in an effort to identify the smallest subset that could reliably classify patients using the SVM model. Briefly, 100 random training sets were selected from the original dataset A, B, and C, each consisting of 80% of the total samples. SVM models were trained for each random dataset and tested on the remaining 20%. For each of the 770 genes a weight score was calculated at each iteration and such scores were summed in order to obtain a cumulative value. These values were sorted in descending order and ROC curves were calculated for the AD classifications using the genes featuring the 5 highest cumulative scores. This analysis was repeated using gene sets spanning between 6 and 100 of the top scorers. The same approach was used to select the genes with the highest weight for the classification of PD patients. A steady maximum level of AUC was observed when the first 47 of the top scoring genes were selected for both the AD and PD classifications, with 13 of those genes being in common between AD and PD. The final signature consisted of a total of 68 genes. This signature set of 68 genes (see Table 1) was used to recalculate the ROC-AUC for the classification analysis. Figure 6 shows the ROC-AUC curves relative to dataset A, B, and C, providing an ROC-$AUC_{AD} = 0.9$, and ROC-$AUC_{PD} = 0.94$, outperforming the 770 gene classifier. Across all 106 samples, two were miss-classified, providing an accuracy using the 68 gene signature of 98%.

Having established the classification algorithm for AD based on cognitive/functional assessment of the subjects we assessed how this correlated to Aβ PET scores to begin assessing whether the TempO-SeqDBS AD classifier might be definitive. Because Aβ PET scans were not reimbursable until after the samples were collected (reimbursement was approved October 2023), the set of patients for which scans were available was limited. There were Aβ PET scores available for 12 of the patients clinically diagnosed with AD. Nine patients were scored as positive for plaques in excess of those typical for their age, supporting a diagnosis of AD, while three patients were scored as negative, not significantly different from the degree of plaque formation for a person their age. All twelve AD patients, whether with positive or negative Aβ PET scores indicative of AD, were classified as AD by the TempO-SeqDBS test. In addition, one patient diagnosed with PD had both a positive Aβ PET scan and DaTscan (used to distinguish between

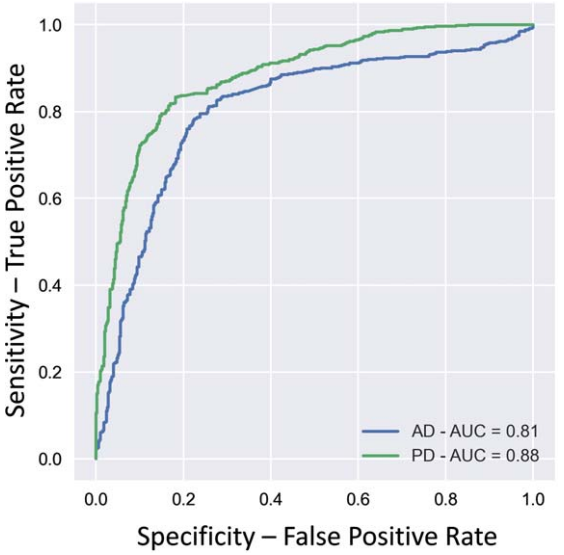

Fig. 4. ROC Curve for classification of Cohort A samples as AD and PD. The ROC curves are shown for the classification of patient samples as AD or PD using the Support Vector Machine (SVM) model (kernel = linear, gamma = 0.1, c = 0.001) and the initially identified 770 gene signature. Use of this model resulted in the best performance for the classification of both AD (ROC-AUC_{AD} = 0.81), and for PD (ROC-AUC_{PD} = 0.86). The model was trained using 80% each of the control, AD, and PD samples in cohort A. As a test for classification accuracy, the trained model was used to classify the remaining 20% of samples. The resulting ROC-AUC vales were 0.81 for AD and 0.88 for PD.

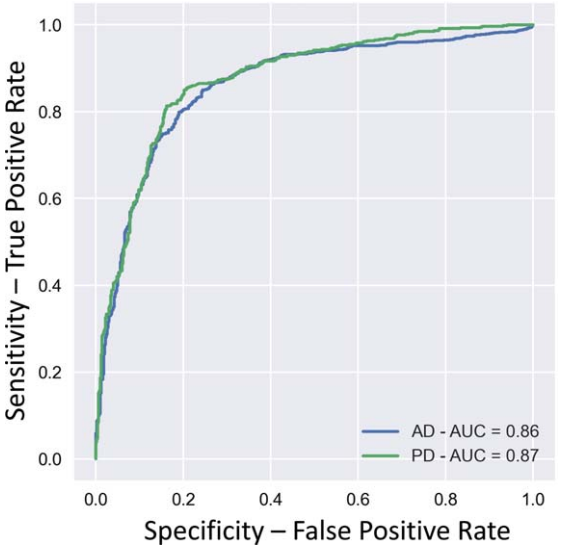

Fig. 5. ROC Curve for Classification of the combined Cohort A and B samples. The ROC curves for the classification of patient samples using the Support Vector Machine (SVM) model (kernel = linear, gamma = 0.1, c = 0.001) and the initially identified 770 gene signature are depicted after pooling together Cohorts A and B to provide a sample set of 106 samples, training with 80% of the samples, and testing with the remaining 20%. For this combined cohort, the ROC-AUC was 0.86 for Ad and 0.87 for PD.

PD where it is positive for loss of dopamine, and essential tremor), and was classified as PD by the TempO-SeqDBS assay.

DISCUSSION

These TempO-SeqDBS test results demonstrate not only the feasibility of identifying gene expression signatures from fingerstick whole blood spotted on filter paper, but also signatures for AD and PD that enabled patients to be classified as AD or PD. These data indicate the feasibility of implementing a test for AD (and PD) that uses a minimally invasive sample that can not only be collected in any doctor's office or clinic, but also can be self-collected, and thus, address health disparities, whether a definitive test or a screen that identifies patients who should be seen for further assessment.

The ROC-AUC values of 0.9 and 0.94 for AD and PD, respectively (Fig. 6) for the combined Cohorts A, B, and C, were excellent, indicating strong putatively diagnostic performance. While we acknowledge the risk of overfitting data from small sample sizes, the

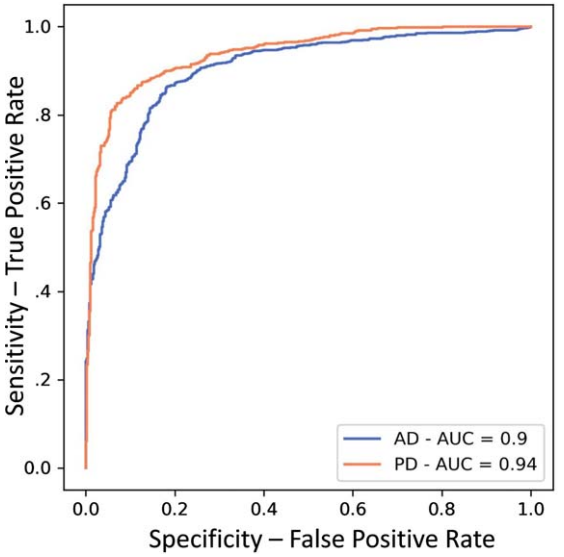

Fig. 6. ROC Curve for classification of Cohort A, B, and C samples as AD and PD, 68 gene signature. The ROC curves are shown for the classification of patient samples as AD or PD using the Support Vector Machine (SVM) model (kernel = linear, gamma = 0.1, c = 0.001) and minimal 68 gene signature. The model was trained using 80% each of the control, AD, and PD samples in cohorts A, B, and C. As a test for classification accuracy, the trained model was used to classify the remaining 20% of samples. The resulting ROC-AUC values were 0.9 for AD and 0.94 for PD.

consistent results for the additional independent AD patient Cohort B of 22 samples, tested separately after retraining the 770 gene signature with all the Cohort A samples, strengthened the validity and generalizability of our findings. Two of the Cohort A, B, and C AD samples were classified as PD, resulting in 98% accuracy calling AD. Thus, the assay provides highly accurate classification whether based on the AUC of 0.9 for AD (Fig. 6) or this final measure of 98% accuracy. It also accurately classified PD (AUC = 0.94, Fig. 6), but the number of PD samples was more limiting, and this will be the focus of a follow-up investigation.

Demonstrating that patients with a clinical diagnosis of AD can be classified from fingerstick blood using a gene expression assay represents the first step towards developing a screening test that informs subjects to see a neurologist for diagnosis earlier then they might otherwise have sought diagnosis, and, potentially, a definitive diagnostic test. With the recent lifting of the curb on reimbursement for Aβ PET scans for AD patients (October 2023), a positive Aβ PET scan compared to non-AD individuals of the same age may also be required for definitive diagnosis. However, in the meantime, Aβ PET scans serve as a useful benchmark test, though itself not definitive.

While Aβ PET scores were only available for 12 (24%) of the 50 patients clinically diagnosed with AD during the sample collection period (as patients had to pay for these scans themselves), 9 were Aβ PET score positive, 3 negative. That all 9 patients with positive Aβ PET scores were also classified as AD by the TempO-SeqDBS test suggests that this classifier may be specific for AD compared to other dementias. That it is not just correlated to positive Aβ PET scores is supported by the observation one patient clinically diagnosed as PD had both a positive Aβ PET scan and positive DaTscan, and this patient was classified by the TempO-DeqDBS test as PD, not AD. A much larger cohort of AD and PD samples with associated Aβ PET scores and DaTscan data is necessary to confirm this conclusion.

Additionally, the classification of all three patients clinically diagnosed with AD, but with Aβ PET scores that were not indicative of AD, requires further investigation. The research demonstrating a blood-based gene expression assay's ability to classify AD patients as much as two years before diagnosis, supports the possibility that the TempO-SeqDBS test can identify patients with AD before they become Aβ PET positive.[21] To determine how early in the progression of AD the TempO-SeqDBS test can classify patients, longitudinal samples need to be collected and tested from individuals before they develop dementia (both those with and without MCI). This will allow us to determine whether classification can occur before patients are biomarker or Aβ positive, and whether all patients classified as AD by the TempO-SeqDBS test will eventually become biomarker and Aβ PET positive.

While we do not have data addressing whether changes in gene expression were connected to disease progression or to disease response, the three most significantly differentially expressed pathways in AD blood relate to the involvement of neutrophils in the immune and inflammatory response, consistent with the literature.[32–35] Among other significant pathways were the cytokine-mediated signaling pathway, cellular response to cytokine stimulus, antigen, and receptor-mediated signaling pathway, all consistent with AD immune cell function literature.[36,37] Whole blood single-cell gene expression studies are planned to address which cells contribute to the classifier, and to pursue the association of specific cells and AD.

Once validated, there are several scenarios for use of the TempO-SeqDBS test, particularly because there is a shortage of neurologists with a wide diversity in access based on geographic location of a patient, which reduces access to care, increases health disparities, and worsens patient outcomes.[11,12] One potential application of the TempO-SeqDBS AD test is as a screening tool for patients who present to their general practitioner (GP) with concerns about early-stage AD or PD. The test results could then be used to identify individuals who require further evaluation. Another potential use is for self-testing, which could directly reduce health disparities. This would also help identify individuals who should seek further evaluation. Thus, without having established that the test is definitive, but rather using the TempO-SeqDBS AD test for screening, neurologists would see patients earlier in the course of their disease, allowing for earlier intervention and potentially greater benefit from therapy. With the recent approval of donanemab and lecanemab for the treatment of early-stage AD, earlier diagnosis can lead to significant benefits for patients. The TempO-SeqDBS AD test could help patients realize these benefits by eliminating the delay caused by limited access to medical specialists, even if the diagnosis itself cannot be made any sooner than with traditional cognitive testing and biomarker analysis conducted by a neurologist. There is also the poten-

tial that the TempO-SeqDBS test could be definitive. To reach this conclusion it will be necessary to correlate the classification by the TempO-SeqDBS assay to biomarker tests and Aβ PET scan data, demonstrating that when classified as AD those patients either are, or over time become, biomarker and Aβ PET scan positive.

AUTHOR CONTRIBUTIONS

Bruce E. Seligmann (Conceptualization; Funding acquisition; Project administration; Supervision; Visualization; Writing – original draft; Writing – review & editing); Salvatore Camiolo (Data curation; Formal analysis; Software; Validation; Writing – review & editing); Monica Hernandez (Investigation; Methodology; Writing – review & editing); Joanne M. Yeakley (Methodology; Writing – review & editing); Gregory Sahagian (Resources; Visualization; Writing – review & editing); Joel McComb (Project administration; Resources; Writing – review & editing).

ACKNOWLEDGMENTS

We thank Joelle McComb, Gail Ramirez and April Tenorio for accruing subjects and obtaining the blood samples.

FUNDING

Funding for this work was in part provided by the National Institute of Health National Institute on Aging grant 1R43 AG065039.

CONFLICT OF INTEREST

Bruce Seligmann, Salvatore Camiolo, Monica Hernandez, Joanne Yeakley, and Joel McComb are current employees of BioSpyder Technologies, Inc., which has commercialized the TempO-SeqDBS assay and may commercialize a test for Alzheimer's or other dementias based on this platform and the data presented. Greg Sahagian is CEO of the Neurology Center of Southern California which received funding to provide samples.

DATA AVAILABILITY

The data supporting the findings of this study are openly available in ENA (https://www.ebi.ac.uk/ena/) with the study accession number PRJEB71651.

REFERENCES

1. World Health Organization, Dementia - WHO 2022 https://www.who.int/news-room/fact-sheets/detail/dementia (2022, accessed 1 September 2022).
2. Gauthier G, Rosa-Neto P, Aorais JJ, et al. *World Alzheimer Report 2021: Journey through the diagnosis of dementia*. London: Alzheimer's Disease International, 2021.
3. Alzheimer's disease facts and figures. *Alzheimers Dement* 2022; 18: 700–789.
4. Dubois B, Villain M, and Frisoni GB. Clinical diagnosis of Alzheimer's disease: recommendations of the international working group. *Lancet Neurol* 2021; 20: 484–496.
5. Thijssen EH, Verberk IMW, Kindermans J, et al. Differential diagnostic performance of a panel of plasma biomarkers for different types of dementia. *Alzheimers Dement (Amst)* 2022; 14: e12285.
6. Hansson O, Seibyl J, Stomrud E, et al. CSF biomarkers of Alzheimer's disease concord with amyloid-b PET and predict, clinical progression: A study of fully automated immunoassays in bioFINDER and ADNI cohorts. *Alzheimers Dement* 2018; 12: 1470–1481.
7. Gobom J, Parnetti L, Rosa-Neto P, et al. Validation of the LUMIPULSE automated immunoassay for the measurement of core AD biomarkers in cerebrospinal fluid. *Clin Chem Lab Med* 2022; 60: 207–219.
8. Jiao B, Liu H, Guo L, et al. Performance of plasma amyloid β, total tau, and neurofilament light chain in the identification of probable Alzheimer's disease in south China. *Front Aging Neurosci* 2021; 13: 749649.
9. Shaw LM, Hansson O, Manuilova E, et al. Method comparison study of the Elecsys® b-amyloid (1-42) CSF assay versus comparator assays and LC-MS/MS. *Clin Biochemistry* 2019; 72: 7–14.
10. Willemse EAJ, Van Maurik IS,Tijms BM, et al. Diagnostic performance of Elecsys immunoassays for cerebrospinal fluid Alzheimer's disease biomarkers in a nonacademic, multicenter memory clinic cohort: the ABIDE project. *Alzheimers Dement (Amst)* 2018; 10: 563–572.
11. Majersik JJ, Ahmed A, Chen I-H A, et al. A shortage of neurologists – we must act now: A report from the AAN 2019 transforming leaders program. *Neurology* 2021; 96: 1122–1134.
12. Lin CC, Callaghan BV, Burke JF, et al. Geographic variation in neurologist density and neurologic care in the United States. *Neurology* 2021; 96: e309–e321.
13. Li H, Honf G, Lin M, et al. Identification of molecular alterations in leukocytes form gene expression profiles of peripheral whole blood of Alzheimer's disease. *Sci Rep* 2017; 7: 14027.
14. Tang R and Liu H. Identification of temporal characteristic networks of peripheral blood changes in Alzheimer's disease based on weighted gene co-expression network analysis. *Frontiers Aging Neurosci* 2019; 11: 83.
15. Rahman R, Islam T, Zaman T, et al. Identification of molecular signatures and pathways to identify novel therapeutic targets in Alzheimer's disease: Insights from a systems biomedicine perspective. *Genomics* 2020; 112: 1290–1299.
16. Patel H, Iniesta R, Stahl D, et al. Working towards a blood-derived gene expression biomarker specific for Alzheimer's disease. *J Alzheimers Dis* 2020; 74: 545–561.

17. Han G, Wang J, Seng F, et al. Characteristic transformation of blood transcriptome in Alzheimer's disease. *J Alzheimers Dis* 2013; 35: 373–386.
18. Rye PD, Booig BB, Grave G, et al. A novel blood test for the early detection of Alzheimer's disease. *J Alzheimers Dis* 2011; 23: 121–129.
19. Voyle N, Keohane A, Newhouse S, et al. A pathway based classification method for analyzing gene expression for Alzheimer's disease diagnosis. *J Alzheimers Dis* 2016; 49: 659–669.
20. Milanesi E, Dobre JM, Cucos CA, et al. Whole blood expression patterns of inflammation and redox genes in mild Alzheimer's disease. *J Inflammation Res* 2021; 14: 6085–6102.
21. Roed L, Grave G, Lindahl T, et al. Prediction of mild cognitive impairment that evolves into Alzheimer's disease dementia within two years using a gene expression signature in blood: A pilot study. *J Alzheimers Dis* 2013; 35: 611–621.
22. Lunnon K, Ibrahim Z, Proitsi P, et al. Mitochondrial dysfunction and immune activation are detectable in early Alzheimer's disease blood. *J Alzheimers Dis* 2012; 30: 685–710.
23. Lee T and Lee H. Prediction of Alzheimer's disease using blood gene expression data. *Sci Rep* 2020; 10: 3485.
24. Shigemizu D, Mori T, Akiyama S, et al. Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through RNA sequencing analysis. *Alzheimer Res Ther* 2020; 12: 87.
25. Leandro GS, Evangelista AF, Lobo RR, et al. Changes in expression profiles revealed by transcriptomic analysis in peripheral blood mononuclear cells of Alzheimer's disease patients. *J Alzheimers Dis* 2018; 66: 1483–1495.
26. Santiago JA, Bottero V and Potashkin JA. Evaluation of RNA blood biomarkers in the Parkinson's disease biomarkers program. *Frontiers Aging Neurosci* 2018; 10: 157.
27. Gao A. Identification of blood-based biomarkers for early stage Parkinson's disease. *medRxiv* 2020; doi: https://doi.org/10.1101/2020.10.22.20217893 [Preprint]. Posted October 27, 2020.
28. Henderson AR, Wang A, Meechoovet B, et al. DNA methylation and expression profiles of whole blood in Parkinson's disease. *Front Genet* 2021; 12: 640266.
29. Yeakley JM, Shepard PJ, Goyena DE, et al. A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. *PLoS One* 2017; 12: e0178302.
30. McDade TW, Ross K, Fried R, et al. Genome-wide profiling of RNA from dried blood spots: convergence with bioinformatic results derived from whole venous blood and peripheral blood mononuclear cells. *Biodemography Soc Biol* 2016; 62: 182–197.