EXPERT REVIEW OF MOLECULAR DIAGNOSTICS
https://doi.org/10.1080/14737159.2019.1626719

Taylor & Francis
Taylor & Francis Group

REVIEW

Blood-based diagnostics of Alzheimer's disease

Gregory Penner[a], Soizic Lecocq[a], Anaëlle Chopin[a], Ximena Vedoya[a], Simone Lista[b,c,d,e], Andrea Vergallo[b,c,d,e], Francois-Xavier Lejeune[d] and Harald Hampel[b,c,d,e]

Q1 [a]NeoNeuro SAS, Villejuif, France; [b]AXA Research Fund & Sorbonne University Chair, Paris, France; [c]Institute of Memory and Alzheimer's Disease (IM2A), Department of Neurology, Pitié-Salpêtrière Hospital, AP-HP, Boulevard de l'hôpital, Paris, France; [d]Brain & Spine Institute (ICM), INSERM U 1127, CNRS UMR 7225, Boulevard de l'hôpital, Paris, France; [e]Sorbonne University, GRC n° 21, Alzheimer Precision Medicine (APM), AP-HP, Pitié-Salpêtrière Hospital, Boulevard de l'hôpital, Paris, France

ABSTRACT

Introduction: This review is focused on the methods used for biomarker discovery for Alzheimer's disease (AD) in blood rather than on the nature of the biomarkers themselves.

Areas covered: All biomarker discovery programs explicitly rely on contrasts in phenotype as a basis for defining differences. In this review, we explore the basis of contrasting choices as a function of the type of biomarker (genetic, protein, metabolite, non-coding RNA, or pathogenic epitope). We also provide an overview of the capacity to identify pathogenic epitopes with our new platform called Aptamarkers. It is suggested that a pre-existing hypothesis regarding the pathophysiology of the disease can act as a constraint to the development of biomarkers.

Expert opinion: Limiting putative biomarkers to those that have a postulated role in the pathophysiology of disease imposes an unrealistic constraint on biomarker development. The understanding of Alzheimer's disease would be accelerated by agnostic, non-hypothesis-based biomarker discovery methods. There is a need for more complex contrasts and more complex mathematical models.

ARTICLE HISTORY
Received 10 January 2019
Accepted 30 May 2019

KEYWORDS
Alzheimer's disease; blood-based biomarkers; Aptamarkers; Aptamers; ; ; ; ;

Q2

1. Overview

Q4 This review characterizes the relationship between the processes used to characterize biomarkers and the type of biomarker platform utilized by using efforts to develop plasma-based biomarkers for Alzheimer's disease (AD) as an example. The paper is organized with a general discussion of how our definition of the disease affects or constrains our efforts to develop biomarkers in general. We will then review how technical limitations to experimental design constrain discovery within genomics, metabolomics, proteomics, and non-coding RNA analysis. We will also present in this context a review of a novel platform developed by the authors (Aptamarkers) and how this platform overcomes many of these constraints. Finally, we will generalize from the AD example to the future of biomarker discovery across pathologies.

2. Relationship between definition of disease and biomarker development (Ex/Alzheimer's disease)

By definition, the level of a biomarker is an indication of a pathological state. Pools of plasma from individuals with a defined state of a pathology versus a healthy state are generally used for initial characterization of biomarkers. The first requirement, therefore, is a clear definition of the disease.

Alzheimer's disease provides a good example of how evolving understanding of the disease affects biomarker discovery. The physical symptoms of a severe loss of memory as a function of aging have been described for as long as people have written records. The designation of an aspect of this as a pathology known as Alzheimer's disease is generally credited to Dr. Alois Alzheimer [1] based on his description of a patient and subsequent analysis of brain tissue with silver staining. Given however that this particular patient was 51 years old at the time the disease presented itself, this is not what we would classically refer to as Alzheimer's disease (AD) now. A clear definition of the disease must include that this is a pathology that affects the aged (generally considered over the age of 65), and the symptoms include both a decline in cognitive capacity and behavioral disturbances. The presence of both Aβ plaques and hyperphosphorylated tau tangles in neural tissue are also necessary for the disease to be termed AD.

The most recent working definition of the disease is referred to as the ATN description [2]. In this understanding the disease is initiated by the deposition of Aβ plaques (A), this is followed by the accumulation of tau tangles (T) and eventually by neurological degradation (N). Biomarker identification in plasma has focused on the characterization of proteins that are postulated to have a role in the pathophysiology of the disease for a specific stage. Thus, there has been a focus on the characterization of Aβ peptides for stage A, tau for stage T, and markers of neurodegeneration such as nfl for stage N.

There are several possible constraints with this approach.

CONTACT Gregory Penner gpenner@neoneuro-aptamers.com NeoNeuro SAS, Villejuif Bio Park, 1 mail du Professeur Georges Mathé, Villejuif 94 800, France

2.1. Mis-aligned individuals

A difficulty with the application of this definition is that there are individuals that exhibit the A and T but do not develop N. Quantitatively, there can also be a lack of alignment between the rate of Aβ plaque accumulation and the onset of tau tangles or neurodegeneration. At least two broad explanations are possible.

(1) The progression from A to T to N is not strictly causal.
(2) Individuals vary in their tolerance to the accumulation of A.

These explanations are not mutually exclusive.

As such, we propose the following definition of the disease that encompasses these possibilities.

'*A subsequent cascade of the disease is triggered when one element of the disease exceeds individual resistance to the previous cascade event*.'

It should be noted that this definition does not limit the cascade events to ATN. The strength of this definition is that it provides room for the exploration of individual variation in resistance to each cascade event. In effect, this definition adds a dimension to our capacity to understand the disease.

3. Greater complexity in contrasts used to discover biomarkers

There is a tendency within the biomarker field to discover biomarkers that delineate a difference between pathological and healthy states, and then to characterize the evolution of the biomarker across different stages of the pathology. This reduces the possibility of identifying biomarkers that are only useful for delineating the difference between different stages of the pathology and not for differentiating the pathology from a healthy state. This may seem counter-intuitive, if a biomarker does not distinguish healthy from affected by the pathology how can it be useful? The key here is the possibility of epistatic interaction among biomarkers. The relationship between an individual biomarker and a pathological state may not be statistically significant, while the interaction between two biomarkers may be highly significant.

Hypothetical example:

When biomarker A is present in a low state and biomarker B in a high state, then the disease progresses more rapidly, whereas a higher abundance of biomarker A and a low abundance of biomarker B are linked to slower disease progression. The level of A and the level of B by themselves are not diagnostic of the disease from a healthy state.

Clearly, this constraint also applies to concept 1. To discover biomarkers that are useful in predicting the level of individual resistance to a stage of a disease will require the initial use of contrasts between individuals with high levels of brain amyloid that do not progress to significant levels of tau tangles and individuals with high levels of brain amyloid that progress normally to subsequent tau tangle development.

4. Brain–blood barrier

Alzheimer's disease and all neurological diseases occur within a tissue that is separated from blood by the blood–brain barrier (BBB). The BBB acts as a biased filter of metabolic events within the brain. Proteins and peptides are cleaved prior to clearance through the BBB. Metabolites pass through at differential rates either passively or through ported systems. In addition, blood should not be considered as a passive carrier of the filtered products. Components in blood are actively metabolizing all molecules that are present, at varying rates into varying products. For example, metastasis of prostate cancer has been shown to inhibit non-trypsin proteolytic cleavage of serum albumin [3], tumor development in breast cancer patients is associated with increased proteolytic activity in blood [4]. Thus, the most useful way to think about biomarkers in blood is to think of them as a shadow of events that are occurring in the brain. Some of the target molecules may be present intact, others may occur in a different form than they do in the brain, and still others may stimulate metabolic pathways in blood such as immune responses that are not active or not as active in the brain. Altogether, all of these factors should be considered as a fingerprint in blood of something occurring on the other side of a barrier.

An unnecessary constraint is imposed if we only consider as targets for blood-based biomarkers molecules that are linked to a suspected pathophysiological process of AD in the brain. This ignores the reality that the BBB is a biased filter, that blood is alive and processing molecules at varying rates, that our understanding of the disease is far from complete, and that an element of the disease may be related to individual resistance at many levels.

5. Summary

It is clear that while measurement of biomarkers in blood such as Aβ peptide ratios and phosphorylated forms of tau protein are useful in diagnosis, their characterization alone cannot be sufficient. There is a need to expand biomarker discovery beyond targets that are linked to pathophysiology in the brain. To overcome this constraint there is a need to change the nature of biomarker discovery.

6. Review of technology platforms

The development of agnostic biomarker approaches has shown great promise with genomics, proteomics, metabolomics, and non-coding RNA characterization. We will review each of these biomarker platforms in terms of how technical constraints reduce their agnostic power and thus their power as discovery tools.

All of these approaches attempt to be less hypothesis driven than approaches aimed at validating a biomarker with a hypothesized pathophysiological effect. In general, these approaches follow the following steps towards the creation of a hypothesis.

a. Characterize variation in a set of molecules in blood (variables)
b. Correlate this variation with variation in the pathology (variates).
c. Build a mathematical model that best describes the way the variables describe variation in the variates.
d. Test the validity of the mathematical model on more samples.

The hypothesis that arises from this process is not that one of these variables explains a portion of the variation in the pathology. The hypothesis that arises is that considered as a group of variables a certain portion of the variability in the pathology can be explained. This represents a difference in the meaning of the biological relationship between the biomarkers and the disease. The shift to explanation of variability enables the inclusion of patterns of biomarkers across individuals that may be contradictory if considered strictly in terms of covariance.

In this review, we will focus on the relationship between the formulation of the experiment and the outcome in regard to the characterization of blood-based biomarkers for AD. The formulation of an experiment includes constraints that are implicit in how different types of molecules are measured and the contrasts that are used to determine meaning. These constraints affect the level to which biomarker discovery can be performed in the absence of a hypothesis.

7. Genomic biomarkers

In genomics characterization of the variables involves defining sequence variants within genes that affect protein sequence. The discovery of sequence variation is derived either from full genome sequencing or the characterization of expressed RNA. The advantage of analyzing RNA is that this substrate does not include non-coding sequences. The disadvantage of expression analysis is that the copy number of the RNA species varies by many orders of magnitude and low abundance sequences are difficult to reliably characterize. Variants of gene are characterized by the variant nucleotide (single nucleotide polymorphism, SNP). Individual SNPs can be assessed based on PCR amplification of the relevant gene, and hybridization of the amplicons to microarray chips.

The Genome-Wide Association Study (GWAS) catalog (https://www.ebi.ac.uk/gwas/efotraits/EFO_0000249) currently contains 94 studies on Alzheimer's disease. Criteria for inclusion in this catalog require that at least 100,000 SNPs have been analyzed. This criteria ensure that each study satisfies the requirement that the analysis of variables is not hypothesis based.

Freudenberg-Hua [5] have provided an excellent review of recent findings in genomic research as a basis for AD. Genetic analysis as interpreted by this review was limited to predicting whether people would become afflicted with the disease or not. As such, information regarding individual rates of disease progression are lost. This is key to applying the understanding derived from this analysis on an individual basis.

On April 6th, 2017, the company 23andme received FDA approval for the diagnosis of ApoE allele information with clients in regard to the increased risk associated with the ApoE-4 allele. The current focus of researchers and companies working on the genetic basis of Alzheimer's disease has shifted from the effect of single alleles by themselves to the consideration of patterns of alleles in terms of overall risk of disease onset or of brain amyloid status [6,7]. Mathematical models based on risk association patterns are subjected to machine learning processes to develop individualized risk assessments.

There are several strengths of this type of approach including coverage of all genetic variation, the ability to trace genetic variance to identified genes and function, and the capacity to individualize risk assessments. A key drawback is obviously that genetic information is a blueprint, human biology is a function of this blueprint along with the effect of the environment throughout the life of each individual. The interactive potential of genetic variants with different environmental influences is still poorly understood in human genetics compared to long history of research in this area in crop genetics.

In crop breeding, the potential performance of genetic variants across varying environments (known as G x E effects) is a more critical consideration than the per se linkage of a gene to an effect within a given environment. To pursue G x E interactions within GWAS analysis would require classification or quantification of the effect of environment over the lifetimes of the individuals prior to their phenotypic assessment. Such an effort is clearly not trivial but could prove extremely useful in providing further refinement to existing prediction models. The gap between a blueprint and the execution of the blueprint may increase over time, classification and incorporation of environmental influences on the reiterative process of blueprint execution may help close this gap. Another key difference between large-scale genetics efforts in crop breeding and human genetics is the continued reliance in crop breeding on a concept referred to as genetic disassociation. Since the human genome was sequenced we tend to think of the genome as a collection of genes, and the variation among individuals as being related to the allelic version of the gene that they have. In reality, the existence of these genes in linkage groups is of biological importance. The average rate of recombination during meiosis is 1.16 per chromosome arm [8]. Given that humans have approximately 25,000 genes this means that the average linkage block contains 469 genes. Clearly, meiosis is more complex than this with the frequency of recombination not being equally probable throughout the genome. However, in crop genetics, the site of recombination in relation to linkage blocks of genes has been highly useful in terms of characterizing variation. It is possible that such analysis would be useful in terms of increasing our understanding of the relationship between genetics and the onset of AD. The full power of genetics is realized through a consideration of the inheritance of linkage blocks, and consideration of the disruption of these linkage blocks on an individual basis. An exception to this was a study by Poduslo et al., that did consider haplotypes. This

study was limited in discovery to two ancestries within two families [9].

A cross-functional application is the discovery of protein biomarkers based on the discovery of a novel allele. One such example for AD is the discovery of the ALZheimer ASsociated protein (ALZAS), based on the identification of a novel allele of the APP gene [10]. This protein has been shown to be enriched in blood of individuals that carry this allele and are affected by AD [11].

7.1. Non-coding RNA biomarkers

Non-coding RNA includes four types of RNA species: micro (miRNA), long noncoding (lnc RNA), circular (circRNA), and piwi-interacting RNA (piRNA). All of these species have been shown to play a role in regulating gene transcription. Biomarker development efforts for AD have focused on the characterization of those noncoding RNA species that are involved in the regulation of genes that are hypothesized to be responsible for the onset of the pathology rather than broad non-hypothesis-based association analysis [12]. A large number of each of these types of noncoding RNA species have been implicated as having a role in the pathogenesis of AD.

A key reason that biomarker development programs with non-coding RNA have chosen to focus on hypothesis-driven discovery strategies focused on biochemical and genetic pathways thought to be involved in the pathophysiology of AD has been the difficulty associated with broad-scale identification of noncoding RNA. The actual cloning and sequencing of noncoding RNA sequences is not trivial. These are predominantly short sequences without defined sequences on either end. Specialized cloning systems have evolved and are commercially available for miRNA, but from personal experience, their use is technically demanding. Just cloning a short sequence of RNA does not mean that it is a meaningful noncoding sequence. It is probable that a cloned library will be contaminated by fragments of coding mRNA. Even if it can be shown that the sequence does not have a complement in the human genome, it does not necessarily mean that it is noncoding. It may have arisen from normal gene expression from one of the many forms of biota that co-exist within us. This is a problem that existing genome sequencing of the human microbiosphere has not proceeded far enough to overcome. To overcome this problem, rules have been developed to define what a particular class of noncoding RNA looks like including definitions of characteristic secondary or tertiary structures. The ultimate identification of a noncoding RNA generally lies in demonstration that it is involved in a gene expression pathway. This is demonstrated by increasing or decreasing the expression of a given sequence and characterizing the response.

Given the amount of effort required to define a noncoding RNA sequence, broad scale, agnostic biomarker discovery efforts with this type of biomarker have not been practical to date for AD. It is increasingly clear that noncoding RNA serves as integrated and networked feedback loops on gene expression. All systems within the human body require complex levels of governance and control. There is a need for systems to act in a manner similar to the mathematical concept of an attractor. Disruption of the system is inevitable as a myriad of external factors at varying levels impact the system. Systems must be designed to respond to such disruptions and return to normal. Noncoding RNA species have evolved as a more sophisticated system of dynamic system regulation than would be possible with hard-wired promoters and enhancers. Noncoding RNA allows more nuanced response of genetic expression in response to environmental stimuli. Given the potential, even the inevitability, that feedback loops from systems that are not directly related to known pathophysiological processes of AD may be integrated with feedback loops that are, it would seem that a more agnostic approach to the characterization of noncoding RNA would prove fruitful. The need for targeted approaches to noncoding RNA biomarker development implies that the basis for such searches needs to come from other biomarker efforts. The key other biomarker efforts that would appear appropriate as guidance to maintain relevance to AD would appear to be genetics and metabolomics with their capacity to identify novel biochemical pathways such as glucose metabolism in early stages of the pathology.

7.2. Protein biomarkers

Proteomics can be divided into two types of approaches, one involves the characterization of proteins that have been identified as useful biomarkers in cerebrospinal fluid (CSF) in blood, while the other approach involves the search for proteins that have not previously been associated with AD. The former approach has been successful with the demonstration that the characterization of Aβ peptide ratios in blood plasma or serum can be used to predict the status of amyloid plaques in brain tissue [13–15]. Initial success in this area has been with the use immunoprecipitation of the targeted peptides followed by mass spectrometry analysis [13,14]. This has been followed by the demonstration of similar levels of accuracy with the automated SIMOA system [15]. Several reviews of blood-based protein biomarkers based on these applications have been recently published [16–25] hence we will not reiterate this information here.

An alternative approach to the characterization of protein biomarkers is the use of broad-based discovery tools such as liquid chromatography, double mass spectrometry (LC-MS /MS) or array-based platforms. In the classic LC-MS/MS approach individual proteins are digested by an enzyme such as trypsin that cuts only after arginine and lysine. In fact, proteomics is almost limited to the use of trypsin because no other enzyme works as well. The peptides are separated from each other through chromatography (usually on the basis of surface hydrophobicity) and their individual mass is determined in the first round of mass spectrometric analysis. Then, each peptide is blown apart and the mass of its constituent parts is determined by a second round of mass spectrometry. Programs exist to construct possible sequences of the peptides based on the mass of their constituents. These possible sequences are then screened against the known protein sequences from the human genome in order to validate the correct sequence. The presence of multiple peptides

from the same sequence can thus be used to validate the presence of a protein in the sample.

A key advance in the application of the LC-MS/MS approach was the completion of the human genome sequence. Once all the genes were known, then it was thought originally that all the proteins encoded by the genome were knowable. It has subsequently become apparent that this could not be realized. It has been estimated that for each gene there are on average 100 different protein species. These proteins differ from each other in terms of post-translational modifications (PTM), sequence variation (SNPs) and variation in RNA cleavage and fusion prior to transcription (AS – alternative splicing) [26]. Given that there are approximately 25,000 human genes this means that there are probably at least 2 million different proteins. It follows that there are many forms of proteins that have not yet been characterized in blood. This broad array of possible protein forms in conjunction with the nature of LC-MS/MS analysis means that the solution space for possible proteins contains many redundant solutions. A proteomic result could be interpreted equally well as different proteins or different variants of the same protein. The software used to explore this solution space is continuing to evolve, but the mathematical issue implicit in redundant solutions imposes a barrier that cannot be overcome without the application of other means of protein characterization.

Another key constraint to the agnostic use of proteomics for the discovery of AD biomarkers is the differences among proteins in their concentration in blood. This difference in the concentration of blood proteins can be on the level of nine orders of magnitude in blood plasma. LC-MS/MS analysis becomes swamped by the more abundant proteins. The number of hits with proteomic analysis is roughly equivalent to the concentration of the protein. Thus, this is not an effective method to identify proteins that are present in low abundance.

Alternative approaches to proteomics that overcome the issue of concentration differences have and are continuing to emerge. The concept of a protein array with antibodies against a panel of known proteins has been used effectively to characterize proteins that differ between those affected by AD and healthy individuals [27]. Hye et al. [28], used classic two-dimensional gel electrophoresis with silver staining to characterize several proteins that were either enriched or depleted in the blood of individuals with AD versus healthy individuals. In both of these studies, the authors analyzed the potential association of each protein found separately as a diagnostic for AD rather than considering them as a combined set of biomarkers.

A group from Perkin Elmer applied a novel, high throughput, agnostic analysis to proteins, and what they referred to as the 'fragmentome', that portion of blood that contains protein fragments and peptides. This portion of the proteome should be of considerable interest to researchers in this area given that proteins are generally degraded prior to clearance from the brain through the BBB. These authors were able to develop fingerprints that were highly associated with the disease. A key difficulty associated with their assay approach, however, is the lack of a clear means of moving from discovery of fingerprints to a more efficient analysis of a targeted subset of targets. Their platform implicitly involves analysis of the entire target set in each sample, thus introducing potential for large experimental error. All three of these groups employed proteomic approaches in a non-hypothesis driven manner with contrasts between healthy and AD-affected individuals.

Array-based approaches to protein identification are not well equipped to identify novel protein variants. Antibodies need to be specifically developed against these defined variant sites and demonstrated to enable specific identification. A good example of where this has been driven by discovery in AD has been the development of commercially available monoclonal antibodies for a wide variety of different phosphorylation sites on tau. This means that given current technical constraints proteomics cannot yet be used for deep, non-hypothesis driven biomarker discovery. Many biomarkers currently in use in other pathologies are protein based. The relative dominance of proteins as biomarkers has been largely driven by the existence of rapid, simple characterization with antibodies, especially in an ELISA format. Advances in sequence analysis, including lateral flow analysis of SNPs, have diminished this advantage.

The most pressing need for more biomarkers covering more aspects of AD is in the early stages of the pathology, not once symptoms of cognitive dysfunction are apparent. At these early stages, the BBB is still fully intact, and the diffusion of proteins from brain tissue to blood is severely constrained. Proteins are cleaved into peptides that are sufficiently small to be passively cleared from the brain and CSF. The small size of these peptides and the fact that they are not derived by trypsin digestion make them difficult targets for analysis by LC-MS/MS or antibodies. Small molecules generally need to be conjugated to a larger molecule in order to be sufficiently antigenic for antibody creation. An alternative approach that overcomes this difficulty is the use of aptamers as ligands for selection. Aptamer selection can be performed against small targets in a free state with the FRELEX selection process [29][25].

8. Metabolomic-based biomarkers

Metabolomics involves the characterization of small molecules. It has been estimated that the human metabolome consists of at least 150,000 different molecules [30,31] of which only 100,000 have been identified [32]. In an excellent review by Wilkins and Trushina [33], metabolomics was divided into four subfields based on the type of analysis and the type of molecules being targeted. Untargeted metabolomics involves characterization of variation in the ratios among thousands of different molecules, while targeted metabolomics refers to analysis of the levels of only a few molecules specific to a particular biochemical pathway. Lipidomics relies on special techniques required to established profiles of water-insoluble lipids. Fluxomics involves the use of radioactive tracers to characterize the dynamic change of metabolite levels. In this case, the relative level of the metabolite is less important than the rate at which it is changing.

Graham et al. [13] analyzed the plasma metabolome across 700 individuals in the ADNI cohort using a variety of contrasts, including levels of Aβ42, tau, MRI analysis of brain structures, and cognitive performance. This study resulted in the identification of a limited number of metabolites associated with

each contrast. Overall, metabolomic analysis supports early changes in energy usage as a result of potential impairment of glycolysis in the brain pre-MCI, this is followed by alteration in lipid and polyamine metabolism in association with the shift from MCI to AD[9]. In late stages of AD, there is a shift to amino acids as an energy source with a particular perturbation of arginine metabolism [14].

Clearly, this is a very powerful approach for discovery. It does, however, fall prey to the same constraints as proteomics, the concentration of metabolites will vary greatly within blood, and measurements of quantity while looking at all metabolites will be swamped by those that are present in high concentration. In proteomics, the identity of a protein can be discerned by pulling apart the masses of its constituent parts, amino acids. Metabolites cannot be reconstructed in silico on the basis of the sequence of their constituent parts. Many completely different metabolites will have similar or identical molecular masses. The two primary tools used for metabolomic analysis are nuclear magnetic resonance (NMR) and mass spectrometric analysis. NMR equipment has a high capital cost but can reliably determine ratios among large numbers of metabolites. Mass spectrometric (MS) analysis is more sensitive, at a lower capital cost, but requires extensive preparation and purification of samples in order to be applied reliably. Of the two applications, NMR is more suitable for the generation of useful clinical predictive or diagnostic fingerprints at this point.

The redundancy of the solution space used for characterizing metabolites represents a more significant constraint to metabolites than it does for proteins. An advantage though is the agnostic characterization of metabolite fingerprints, either by NMR or MS analysis. In this approach, the identity of the metabolites is not necessarily known, only the reproducibility of the signal is required. This type of approach has been successfully used to identify bacterial species [34]. As such, this type of analysis would be appropriate for deep, non-hypothesis driven discovery.

8.1. Aptamarkers

Aptamarkers represent a new approach that is broader than the 'omics' approaches reviewed herein. With Aptamarkers we are attempting to characterize epitopes in blood that differ in their level of enrichment between individuals exhibiting a particular pathophysiological phenotype of AD. We achieve this through the use of enriched aptamer libraries. Single-stranded oligonucleotides have the capacity to mimic antibodies and bind specifically and with reasonably high affinity to molecular targets. Aptamer technology consists of the selection of deep libraries of random synthetic oligonucleotide sequences for their capacity to bind to specified targets. We have expanded this application to the use of selected libraries of aptamers for a broad array of unknown targets in blood.

The Aptamarker platform consists of three steps.

(1) Selection of aptamer libraries that are enriched for sequences that bind to molecular targets that are enriched in a pool of blood samples from those affected by the pathology.
(2) Characterization of each of the aptamers in the enriched library following a single round of selection against individual blood samples from individuals varying in the level of pathophysiology presented through next-generation sequencing.
(3) Application of a subset of meaningful Aptamarkers in a single round of selection against blood from a larger set of individuals that vary in the level of pathophysiology presented through qPCR analysis. This step leads to the construction of a mathematical model that provides an overall association of aptamarker frequency and disease state.

We have demonstrated as a proof of principle that this approach can lead to the identification of a group of aptamers with utility for predicting brain amyloid status in cognitively normal individuals in association with the INSIGHT-PreAD study group [35]. This approach is truly agnostic in that we do not know what molecular targets the aptamers we have characterized bind to. We do know the sequences of the aptamers that we are using however, and this means that we can reproduce our analysis across individuals and that this approach can be applied as a diagnostic platform. A strength of the approach is the reliance on simple qPCR analysis for characterization of each diagnostic aptamer.

The Aptamarker platform is not affected by the problem of solution space redundancy that afflicts proteomics and metabolomics. The approach does suffer from some of the contamination issues implicit with non-coding RNA in that not all the aptamers present in an enriched library are necessarily meaningful. The method shares a capacity with genomics in terms of adding a dimension of breadth of analysis, it provides an additional dimension of depth by normalizing the concentrations of the targets as the result of PCR amplification over successive selection rounds and by enabling characterization of changes in states over time. The approach is new, and there is a need to expand validation over more individuals and to apply selections to more contrasts. Theoretically at least, this approach has tremendous promise in enabling deeper non-hypothesis-based discovery than other platforms.

9. Expert opinion

9.1. Contrasts

All biomarker discovery programs require the establishment of a contrast between groups. The most common contrast used is between healthy individuals and those affected by AD. We suggested a working definition of the disease in the introduction, 'A progressive decline in cognitive function where the trajectory of accumulation of Aβ plaques and tau tangles exceeds a personal tolerance of these factors.' This definition results in the reconsideration of individuals with existing biomarker profiles that are not expected based on their pathophysiological status from mis-aligned to being representatives of the top portion of a continuum. As such, a contrast between groups of individuals that all exhibit high levels of brain amyloid but differ in that one group exhibits cognitive dysfunction sooner, while the other group does not would be

useful to delineate an extreme of the continuum. The use of contrasts like this will be useful in recognizing the continuum of the disease and potentially providing a basis for the definition of subtypes within the pathology.

The characterization of biomarkers of all types can be done with measurable levels of confidence and reproducibility. In most cases, the reproducibility of the biomarker measurement is higher than the reproducibility of the phenotypic assessments based on a combination of clinical analysis and brain imaging. Although it is clearly crucial to maintain relevance with phenotypic measurements, it may be useful to begin establishing contrasts for second-generation biomarker studies in AD based on differences in biomarkers. A strength of working with existing cohort samples is that ultimate trajectories through disease states are known. As such, it is possible to establish contrasts based on differences in metabolomic patterns at MCI between pools of individuals that all go on to be affected by AD.

The establishment of contrasts based on differences in biomarkers will be even more fruitful if such approaches are based on different types of biomarkers. As an example, a contrast established based on differences in lipid profiles focused on the characterization of variation in non-coding RNA avoids the tautology of basing a contrast on the same type of biomarker.

9.2. Plethora of known variables

One of the most significant constraints to progress with blood-based AD biomarkers is the emerging picture of this disease as multifaceted. A preliminary study of 100 individuals with any approach requires considerable effort. If we consider the potential effect of ApoE allele, age, and gender alone, we have already partitioned 100 individuals into subgroups that are too small for meaningful statistical analysis.

It is likely more useful at this stage to consider these other known factors as separate variables rather than as classes. This approach requires a move away from direct association between biomarkers and a state of the disease towards the consideration of the capacity of biomarkers to explain variation across individuals for a given disease state. This is dealt with more directly in the next section.

9.3. Association of biomarkers with AD

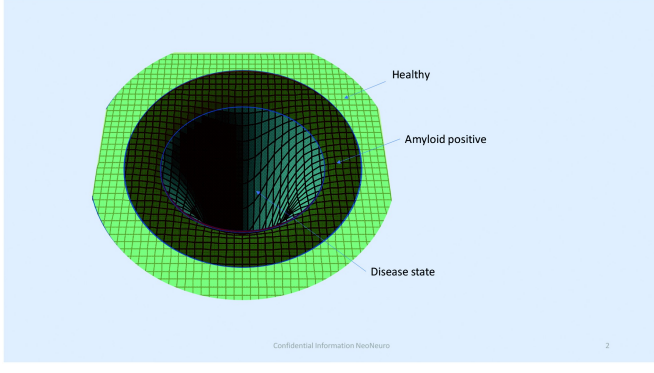
The traditional standard for evaluating the association of a potential biomarker with a trait is to consider covariance. If the level of a biomarker covaries either positively or negatively with the presence of a pathology, then it can be used for diagnostic or predictive purposes. This works well for simple diseases where the pathophysiology is the same for all individuals afflicted. It is probable that we need to consider other approaches when the pathophysiology is as complex as it is with AD and the manifestation of the pathology has the potential for high levels of variability among individuals. Let us consider cardiac arrest as a pathology. An excellent biomarker for the existence of this pathological event is the release of troponin into the blood. This does not, however, provide us with any insight into the cause of the cardiac arrest. One possible cause is a blood clot that leads to heart failure, while another possible cause could be a disruption in the signaling between the brain and the heart. Clearly, these two different causes will have entirely different biomarker fingerprints.

We suggested in a previous section that it was useful to think of AD as analogous to the mathematical concept of an attractor. In this concept, we can visualize the current state as the position of a body in motion on the surface. Perturbations from a healthy state occur constantly. Wellness is similar to a cavity that it is difficult to climb out of. As we age, the walls of this cavity change, it becomes easier for the body in motion to climb higher on the wall. To bring this analogy back to the reality of AD, we consider the AD as a separate attractor that the body in motion can fall into if it climbs too far out of the healthy state (Figure 1). Our genetics lay the basis for the shape of this cavity, and the basis for how this shape changes over time.

In keeping with this conceptualization, the meta-outcome of biomarker development programs in AD can be considered as a mapping of the nature of this surface across individuals. When we consider biomarkers in this light, we are expanding our basis for understanding. There is a tendency to consider AD as a pathology that unfortunately occurs to certain individuals. It may be equally useful to consider a loss of capacity of wellness over time as a risk factor, for which AD is one possible outcome.

Biomarker analysis combined in this manner will lead to the definition of not where an individual is on a common surface, but rather a characterization of individual surfaces and how they differ. Treatments of AD affect the topology of these individual surfaces. Understanding the nature of the effect of a treatment on a topological basis would lead to an improved capacity to predict which surfaces and hence which groups of individuals, a treatment would be effective on.

There is an understandable reluctance to move beyond simple associations between individual biomarkers and a pathology. The more complex the mathematical model becomes the more difficult it becomes to understand what the model is telling us in real biological terms. We do not pretend at this time to have any basis for translating the surface map drawn in Figure 1 to the pathophysiology of the disease. We are merely suggesting that this could be a meaningful mathematical basis for bringing together all of



COLOUR ONLINE BLACK & WHITE IN PRINT

Figure 1. Concept of attractor surface as a description for trajectories leading to Alzheimer's disease.

the biomarker information that is rapidly accumulating, along with brain imaging data, and behavioral analysis.

In the meantime, there is a disease to overcome. Biomarkers are not just a potentially useful means of improving clinical trials and assisting the application of effective treatments. Biomarkers for a disease as complex as AD represent a necessary prerequisite for the development of treatments. The role of biomarkers from different biological systems to characterize variation across the continuum of the pathology is necessary in order to continue to evolve and refine the definition of the disease. There is a pressing need for more contrasts and for more interactive biomarker studies in order to build a broader understanding as a basis for the development of effective treatments for this devasting disease.

9.4. Extension of concepts raised to other pathologies

Alzheimer's disease is not distinct from other pathologies in terms of evolving definition of the disease, evolving understanding of pathophysiology, attractiveness of plasma biomarkers even though this is not the site of the pathology. As such, many of the concepts raised here are just as appropriately raised regarding biomarker development for other pathologies.

9.4.1. Constraints imposed by phenotype characterization

Science, in general, is always constrained by our ability to measure phenotype. The measurement of phenotype is defined by the evolution of the tools that are used to measure it, by clinicians need to provide a narrative to patients, and by the way, we imagine meaningful distinctions across a continuum of observed results. Human height is a quantitative trait that exhibits a continuous distribution. The application of a binary concept such as defining short or tall is implicitly arbitrary. It is important to recognize that the definition of the onset of a disease can also be arbitrary.

Wherever possible, across diseases, it is necessary to correlate biomarker predictive capacity to exactly what is being measured. There is a danger in making the leap beyond characterization of a phenotype this to the prediction of disease based on single biomarkers.

9.4.2. Biological complexity

Biomarker studies are based on contrasts built from cohorts of enrolled individuals. For all pathologies, this necessarily introduces complexity. Each individual enrolled varies not only for the pathology under study but in terms of health for a host of other pathologies. The ability to analyze complexity is no longer constrained by computing capacity, it is constrained by degrees of freedom and sample population size.

There is a need for increased interaction among biomarker researchers across pathologies. The focus in the future will shift from treating a pathology to maintaining health against the threat of all pathologies. This shift needs to be accelerated in terms of design of biomarker discovery.

9.4.3. Blood as more than a passive carrier of biomarkers

The concept that detection of a pathology may be the end result of a cascade of metabolic events in blood triggered by the presence of a pathology is not limited to pathologies that occur on the brain side of the BBB. For this reason, and because of the concepts articulated above, there is a need for agnostic, non-hypothesis based approaches to biomarker discovery across pathologies.

As discussed in this review, platforms such as proteomics, metabolomics, and non-coding RNA have technical constraints that compromise capacity for agnostic application. GWAS and Aptamarkers are less constrained in this way. A combination of these two approaches has considerable promise for enabling significant advances in diagnosing and understanding all diseases.

Funding

The development of this manuscript was funded by NeoNeuro. Q3

Declarations of Interest

NeoNeuro, the company that funded this manuscript, is a private company working to develop an improved blood-based diagnostic system for the prediction of brain amyloid status. NeoNeuro is wholly owned by NeoVentures Biotechnology Inc. (A Canadian registered private company). Gregory Penner and Ximena Vedoya own all shares in NeoVentures Biotechnology Inc. Gregory Penner has also received grant funding from GSK and has ownership over NeoNeuro. Ximena Vedoya has an ownership position over NeoNeuro. Soizic Lecocq and Anaëlle Chopin are employees of NeoNeuro. Simone Lista has received lecture honoraria from Roche. Harald Hampel serves as a Senior Associate Editor for the journal, 'Alzheimer's & Dementia'; has received lecture fees from Biogen, Roche, Eisai; has received research grants from Pfizer, Avid and MSD Avenir (paid to the institution); received travel funding from Functional Neuromodulation, Axovant, Eli Lilly and company, Takeda and Zinfandel, GE-Healthcare and Oryzon Genomics; received consultancy fees from Qynapse, Jung Diagnostics, Cytox Ltd., Axovant, Anavex, Takeda and Zinfandel, GE Healthcare and Oryzon Genomics, and Functional Neuromodulation; and has participated in scientific advisory boards of Functional Neuromodulation, Axovant, Eli Lilly and company, Cytox Ltd., GE Healthcare, Takeda and Zinfandel, Oryzon Genomics and Roche Diagnostics. Harald Hampel is supported by the AXA Research Fund, the 'Foundation partenariale Sorbonne Universite' and the 'Fondation pour la Recherche sur Alzheimer', Paris, France. Ce travail a bénéficie d'une aide de l'Etat 'Investissements d'avenir' ANR-10-IAIHU-06. The research leading to these results has received funding from the program. 'Investissements d'avenir' ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Agence Institut Hospitalo-Universitaire-6). Harald Hampel is co-inventor in the following patents as a scientific expert and has received no royalties:

- In vitro Multiparameter Determination Method for the Diagnosis and early diagnosis of Neurodegenerative Disorders Patent Number: 8,916,388.
- In vitro Procedure for Diagnosis and early Diagnosis of Neurodegenerative Diseases Patent Number: 8,298,784
- Neurodegenerative Markers for Psychiatric Conditions Publication Number: 20,120,196,300
- In vitro Multiparameter Determination Method for the Diagnosis and Early Diagnosis of Neurodegenerative Disorders Publication Number: 20,100,062,463
- In vitro Method for the Diagnosis and Early Diagnosis of Neurodegenerative Disorders Publication number: 20,100,035,286
- In vitro Method for the Diagnosis of Neurodegenerative Diseases Patent Number: 7,547,553

- CSF Diagnostic in Vitro Method for Diagnosis of Dementias and Neuroinflammatory Diseases Publication Number: 20,080,206,797
- In vitro Method for The Diagnosis of Neurodegenerative Diseases Publication Number: 20,080,199,966
- Neurodegenerative Markers for Psychiatric Conditions Publication Number: 20,080,131,921

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewers disclosure

Peer reviewers on this manuscript have no relevant financial relationships or otherwise to disclose.

References

1. Alzheimer A. Über einen eigenartigen schweren Erkrankungsprozeβ der Hirnrincle. Neurol Cent. 1906;25:1134.
2. Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14(4):535–562.
3. Dekker LJ, Burgers PC, Charif H, et al. Differential expression of protease activity in serum samples of prostate carcinoma patients with metastases. Proteomics. 2010 Jun;10(12):2348–2358.
4. Tamkovich S, Bryzgunova O. Protease activity and cell-free DNA in blood plasma of healthy donors and breast cancer patients. J Immunoassay Immunochem. 2016;37(2):141–153.
5. Freudenberg-Hua Y, Li W, Davies P. The role of genetics in advancing precision medicine for Alzheimer's disease-a narrative review. Front Med (Lausanne). 2018 Apr 24;5: 108.
6. Escott-Price V, Shoai M, Pither R, et al. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. Neurobiol Aging. 2016;4:214e7–214e1116.
7. Escott-Price V, Myers AJ, Huentelman M, et al. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. Ann Neurol. 2017;82(2):311–314.
8. Falek and Chiarelli. Meiotic chromosome of man. P19-22 in meiosis research III. New York (NY): MSS Information Corporation; 1972. Melynk, John, Published.
9. Poduslo SE, Huang R, Huang J, et al. Genome screen of late-onset Alzheimer's extended pedigrees identifies TRPC4AP by haplotype analysis. Am J Med Genet B Neuropsychiatr Genet. 2009 Jan 5;150B (1):50–55.
10. Kienzl E, Jellinger K, Janetzky B, et al. A broader horizon of Alzheimer pathogenesis: ALZAS - an early serum biomarker. J Neural Transm Suppl. Vol 109, 2002;87–95.
11. Jellinger KA, Janetzky B, Attems J, et al. Biomarkers for early diagnosis of Alzheimer disease: 'ALZheimer ASsociated gene'–a new blood biomarker? J Cell Mol Med. 2008;12(4):1094–1117.
12. Idda ML, Munk R, Abdelmohsen K, et al. Noncoding RNAs in Alzheimer's disease. Wiley Interdiscip Rev RNA. 2018 Mar;9(2): e1463.
13. Graham SF, Chevallier OP, Elliott CT, et al. Untargeted metabolomic analysis of human plasma indicates differentially affected polyamine and l-arginine metabolism in mild cognitive impairment subjects converting to Alzheimer's disease. PLoS One. 2015.
14. Vural H, Sirin B, Yilmaz N, et al. The role of arginine-nitric oxide pathway in patients with Alzheimer disease. Biol Trace Elem Res. 2009 Summer;129(1–3):58–64.
15. Nakamura A, Kaneko N, Villemagne VL, et al. High performance plasma amyloid-β biomarkers for Alzheimer's disease. Nature. 2018 Feb 8;554(7691):249–254. Epub 2018 Jan 31.
16. Thijssen E, Verberk IMW, Vanderstichele HM, et al. A prototype Simoa assay quantifying plasma amyloid beta 1-42 and 1-40 isoforms can differentiate participants with AD from healthy control subjects. Alzheimer's Dementia. 2018; 14(7):P1039.
17. Ovod V, Ramsey KN, Mawuenyega KG, et al. Amyloid β concentrations and stable isotope labeling kinetics of human plasma specific to central nervous system amyloidosis. Alzheimer's Dementia. 2017. DOI:10.1016/j.jalz.2017.06.2266
18. Molinuevo JL, Ayton S, Batrla R, et al. Current state of Alzheimer's fluid biomarkers. Acta Neuropathol. 2018 Nov 28. DOI:10.1007/s00401-018-1932-x.
19. Hampel H, O'Bryant SE, Molinuevo JL, et al. Blood-based biomarkers for Alzheimer disease: mapping the road to the clinic. Nat Rev Neurol. 2018 Nov;14(11):639–652.
20. O'Bryant SE, Mielke MM, Rissman RA, et al. Biofluid based biomarker professional interest area. Blood-based biomarkers in Alzheimer disease: current state of the science and a novel collaborative paradigm for advancing from discovery to clinic. Alzheimer's Dementia. 2017 Jan;13(1):45–58.
21. Henriksen K, O'Bryant SE, Hampel H, et al. Blood-based biomarker interest group. The future of blood-based biomarkers for Alzheimer's disease. Alzheimer's Dementia. 2014 Jan;10(1):115–131.
22. Lista S, Faltraco F, Prvulovic D, et al. Blood and plasma-based proteomic biomarker research in Alzheimer's disease. Prog Neurobiol. 2013Feb-Mar;101-102:1–17.
23. Hampel H, Frank R, Broich K, et al. Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. Nat Rev Drug Discov. 2010 Jul;9(7):560–574.
24. Blennow K, Hampel H, Weiner M, et al. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat Rev Neurol. 2010 Mar 6;(3): 131–144. DOI:10.1038/nrneurol.2010.4
25. With sudden progress, blood Aβ rivals PET at detecting amyloid; https://www.alzforum.org/news/conference-coverage/sudden-progress-blood-av-rivals-pet-detecting-amyloid, accessed December 10, 2018
26. Ponomarenko, Elena A., Poverennaya, Ekaterina V., Ilgisonis, Ekaterina V., et al. The size of the human proteome: the width and depth. Int J Anal Chem. Volume 2016; 6. Article ID 7436849.
27. Guo L-H, Alexopoulos P, Wagenpfeil S, et al. Plasma proteomics for the identification of Alzheimer disease. Alzheimer Dis Assoc Disord. 2013;27(4):337–342.
28. Hye A, Lynham S, Thambisetty M, et al. Proteome-based plasma biomarkers for Alzheimer's disease. Brain. 2006 November 1;129 (11):3042–3050.
29. Lecocq S., Spinella K., Dubois B., et al. Aptamers as biomarkers for neurological disorders. Proof of concept in transgenic mice. PLoS One. 2018 Jan 5. DOI:10.1371/journal.pone.0190212.
30. Markley JL, Brüschweiler R, Edison AS, et al. The future of NMR-based metabolomics. Curr Opin Biotechnol. 2017 Feb;43:34–40.
31. Lopez MF, Mikulskis A, Kuzdzal S, et al. High-resolution serum proteomic profiling of Alzheimer disease samples reveals disease-specific, carrier-protein-bound mass signatures. Clin Chem. 2005;51(10):1946–1954.
32. Wishart DS, Jewison T, Guo AC, et al. HMDB 3.0 — the human metabolome database in 2013. Nucleic Acids Res. 2013 Jan 1;41 (D1):D801–7.
33. Wilkins JM, Trushina E. Application of metabolomics in Alzheimer's disease. Nucleic Acids Res. 2013 Jan;41(Database issue):D801–7. *Frontiers in Neurology* vol. 8 719. 12 Jan. 2018.
34. Delpassand Ebrahim S, Chari Mohan V, Stager Charles E, et al. Rapid identification of common human pathogens by high resolution proton magnetic resonance spectroscopy. J Clin Microbiol. 1995 May; 33(5):1258–1262.
35. Penner G., Soizic L., Chopin A., et al. Novel use of aptamer libraries for prediction of amyloid status from blood serum. Poster LBP26, JPAD. 2018;5(4):S171–172.