Analysis of the Human Plasma Proteome Using Multi-Nanoparticle Protein Corona for Detection of Alzheimer's Disease

Claudia Corbo,* Andrew A. Li, Hossein Poustchi, Gha Young Lee, Sabrina Stacks, Roberto Molinaro, Philip Ma, Theo Platt, Shahed Behzadi, Robert Langer, Vivek Farias, and Omid C. Farokhzad*

As the population affected by Alzheimer's disease (AD) grows, so does the need for a noninvasive and accurate diagnostic tool. Current research reveals that AD pathogenesis begins as early as decades before clinical symptoms. The unique properties of nanoparticles (NPs) may be exploited to develop noninvasive diagnostics for early detection of AD. After exposure of NPs to biological fluids, the NP surface is altered by an unbiased but selective and reproducible adsorption of biomolecules commonly referred to as the biomolecular corona or protein corona (PC). The discovery that the plasma proteome may be differentially altered during health and disease leads to the concept of disease-specific PCs. Herein, the disease-specific PCs formed around NPs in a multi-NPs platform are employed to successfully identify subtle changes in plasma protein patterns and detect AD (>92% specificity and ≈100% sensitivity). Similar discrimination power is achieved using banked plasma samples from a cohort of patients several years prior to their diagnosis with AD. With the nanoplatform's analytic ability to analyze pathological proteomic changes into a disease-specific identifier, this promising, noninvasive technology with implications for early detection and intervention could benefit not only patients with AD but other diseases as well.

1. Introduction

Alzheimer's disease (AD) is a type of dementia affecting one in ten people over the age of 65[1,2] for which existing diagnostic methods are both very invasive and costly. This neurodegenerative disorder causes major disruptions in cognitive processes such as memory and behavior, with yearly increases in the number of affected individuals worldwide.[2,3] Especially since its prevalence is expected to increase to 88 million by 2050 from 55 million in 2019, the demand for appropriate treatment will continue to grow.[2] Because current research hypothesizes that AD pathogenesis begins 20–30 years before a clinical diagnosis from presenting symptoms,[4–9] the development of robust early diagnosis remains a crucial clinical and social priority.

Current approaches to diagnosis of AD require an invasive lumbar puncture to measure levels of Aβ42, tau, and p-tau in cerebrospinal fluid (CSF).[10,11] Moreover, because no single biomarker can accurately

Dr. C. Corbo, G. Y. Lee, S. Stacks, Dr. S. Behzadi, Prof. O. C. Farokhzad
Center for Nanomedicine
Department of Anesthesiology
Brigham and Women's Hospital
Harvard Medical School
Boston, MA 02115, USA
E-mail: claudia.corbo@unimib.it; ofarokhzad@bwh.harvard.edu

Dr. C. Corbo
School of Medicine and Surgery
Nanomedicine Center Nanomib
University of Milano-Bicocca
Vedano al Lambro 20854, Italy

Dr. A. A. Li
Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213, USA

Dr. H. Poustchi
Digestive Oncology Research Center
Digestive Disease Research Institute
Tehran University of Medical Sciences
Tehran 4117-13135, Iran

Dr. R. Molinaro
Department of Medicine
Brigham and Women's Hospital
Harvard Medical School
77 Avenue Louis Pasteur, Boston, MA 02115, USA

Dr. P. Ma, Dr. T. Platt
Seer, Inc.
3800 Bridge Parkway Redwood City, CA 94065, USA

Prof. R. Langer
Koch Institute for Integrative Cancer Research at MIT
Cambridge, MA 02139-4307, USA

Prof. V. Farias
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02142, USA

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
HEALTHCARE
MATERIALS**

www.advhealthmat.de

diagnose AD,[12,13] patients are subjected to multiple analyses to increase the probability of an accurate diagnosis. Thus, it is crucial to discover more accurate diagnostic markers for the early detection of AD and develop a robust, non-invasive method of measuring them. Analyzing patients' blood (serum and plasma) is an appealing possibility, since drawing blood creates only minor discomfort and allows for frequent testing, follow-up, and access to clinical trials.

Nanoparticles (NPs) in biological fluids interact with a wide range of biomolecules, particularly proteins, which are later adsorbed onto the NPs' surface to form a biomolecular layer referred to as the protein corona (PC).[14–18] The PC gives the NPs a new biological identity distinct from their pristine synthetic identity,[19–21] which is "seen" by cells and processed by biological systems in a specific manner depending on its composition.[17] Indeed, the specific content of the PC dramatically affects the biological fate of the NPs in terms of uptake and interactions with cells, pharmacokinetics, therapeutic efficacy, toxicity, and circulation time.[20,22–31]

It is increasingly being accepted that alterations in the plasma proteome, caused by a spectrum of diseases, can affect the PC composition, generating disease-specific protein coronas.[32–34] Therefore, characterization of personalized PCs may become a tool for the detection of variations in plasma protein concentrations and a valuable alternative method for capturing fluctuations in circulating protein concentrations undetectable with conventional blood analysis. The scientific communities in multiple disciplines have long been working to develop blood-based diagnostic technologies.[35–42] Indeed, recent studies have tested the feasibility of employing the personalized-PC formed on NPs to detect pancreatic cancer or assess the status of acute viral infection using dynamic-light scattering[38,39] and SDS-PAGE.[40] However, these approaches need further development to improve specificity, prediction rate, and sensitivity. Recently, our group and others have employed single- and/or multi-nanoparticle protein coronas to detect cancer and inflammatory diseases.[43–45] Herein, combining complementary expertise in array recognition, bioinformatics, and machine learning, we describe the development of a multi-nanoparticle protein corona nanoplatform capable of detecting important proteins among subtle corona changes and robustly identifying AD in its early stages. The nanoplatform is composed of 6 NPs similar in size, but having different composition and surface functionalization (i.e., 100 nm polystyrene and silica NPs in their plain, -amino, and -carboxyl conjugated forms). The differences in their surface properties mean that the 6 NPs form similar but distinct protein corona patterns and, in turn, the entire set of coronas is different when incubated in plasma from diseased versus healthy individuals, that is, Alzheimer's disease protein coronas (AD-PCs) versus healthy protein coronas (H-PCs). Using quantitative LC-MS/MS analysis, we identified and quantified the proteins constituting the overlapping but distinct protein corona patterns for each plasma sample in this 6-NP combination. We then used the compositions of AD-PCs and H-PCs in machine-learning analyses that generated an AD-specific protein corona identifier for later use in validating the nanoplatform technology on a blinded set of plasma samples. The AD and healthy control samples were divided into a training set for supervised classification and predictor identification and a test set for the prediction of blinded samples, including plasma

samples for the cohort studies (**Figure 1**). The nanoplatform discriminated between AD patients and healthy individuals with an accuracy of ≈94%, and achieved successful early detection of AD with an accuracy of ≈96% in a retrospective cohort of individuals healthy at the time of plasma collection but diagnosed with AD several years later.

2. Results and Discussion

2.1. Nanoparticle Characterization and Protein Corona Formation

The nanoplatform was composed of 6 NPs with different surface modifications intended to form varying protein corona profiles upon incubation in plasma samples. The 6 NPs were 100 nm silica (S) or polystyrene (P) nanoparticles with either plain (P and S), amino-conjugated (P-NH2 and S-NH2), or carboxyl-conjugated (P-COOH and S-COOH). As initial proof-of-concept, the size, zeta potential, and morphology of the NPs with and without PC were examined for differences between AD and control plasmas. The NPs' size distributions were characterized using Nanosight's nanoparticle tracking analysis. Before plasma incubation, all NPs were homogeneous in size (**Figure 2**A and Figure S1, Supporting Information), with the polystyrene NPs ranging from 90 to 100 nm and the silica NPs ranging from 80 to 100 nm. Our characterization confirmed the manufacturer's specifications that the NPs had a negative surface charge. After plasma incubation, however, we found that all 6 NPs coated with protein corona displayed less homogeneity, that is, increased size and wider distribution (Figure S1, Supporting Information; scatter plot). Consistent with the literature, the average size increase was 30 nm, indicative of an average thickness of 15 nm for the protein corona layer.[46,47] The surface charge of the protein-coated NPs also decreased, consistent with the typical charge of plasma proteins of ≈−20 to −30 mV (Figure 2A). The surface charge differences between AD plasma-incubated NPs and healthy plasma-incubated NPs were not significant enough to be a factor in AD discrimination. NP morphology was characterized using transmission electron microscopy (TEM). Incubation in plasma did not alter the spherical and homogenous morphology of the NPs, but a thin layer was visible around the NPs, which was confirmed by the size increase from the characterization described above (Figure 2B).

2.2. Protein Corona Analysis

Proteins constituting the different PCs of NPs were resolved by gel electrophoresis (SDS-PAGE) and later stained for examination with Coomassie Brilliant Blue (**Figure 3**). As expected, gel eletrophoresis patterns of the protein corona were different for the 6 nanoparticles, for example, silica and polystyrene NPs demonstrated different abilities to adsorb proteins. Indeed, we observed less intense bands in the PCs of silica NPs compared to the bands of the polystyrene NPs PCs. Also, the PC of plain polystyrene was much more abundant in the range of molecular weights 55–65 kDa, indicating that albumin was more abundant in this PC than in others. This scenario was observed both for AD patients' plasma and for the control plasma, as highlighted by

**ADVANCED
SCIENCE NEWS**
www.advancedsciencenews.com

**ADVANCED
HEALTHCARE
MATERIALS**
www.advhealthmat.de
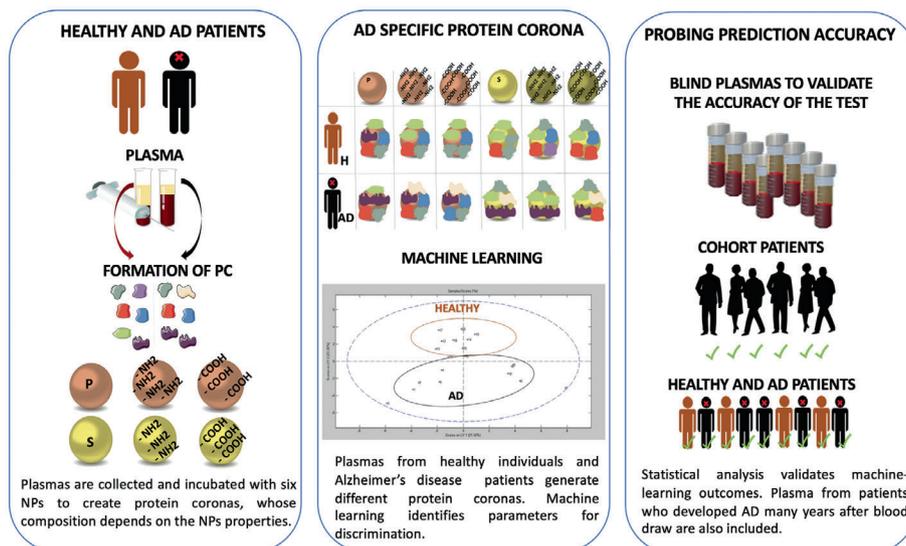
Figure 1. Schematic representation of the personalized PC test workflow. 6 NPs (polystyrene and silica, both in their plain, amino-, and carboxyl-conjugated forms) were incubated with the plasma of healthy individuals and AD patients, and the PCs formed on each NPs were identified and quantified by LC-MS/MS analysis. The personalized PCs formed on the 6 NPs result in the enrichment of an overlapping but distinct pool of specific plasma proteins representing the bases for subsequent statistical analysis. Using classification approaches, we identified the proteins whose contributions to the PC were consistently different between AD and healthy samples. Those proteins were then used to establish criteria for discrimination between the two groups. Next, we evaluated the accuracy of the system by analyzing blind plasmas. We also included a set of cohort samples of patients diagnosed with AD several years after the plasma collection.

the blue arrows in Figure 3. Analysis of the bands in the PCs by densitometry confirmed that silica NPs generally adsorbed lower amounts of proteins, but those proteins were more diverse. Indeed, while more than one band was detected in the range of albumin for polystyrene NPs' PC (red arrows), several differences were observed between the PCs of AD patients and healthy individuals, especially for silica NPs (green arrows). While a detectable difference was reassurance that the different NPs were each creating a different protein corona profile, the gel differences alone were not sufficient to accurately discriminate AD patients from healthy controls. Next, the composition of the PC was analyzed by mass spectrometry to characterize the individual proteins and their relative amounts. We identified 446 and 682 proteins in the non-cohort and cohort data sets, respectively. We also determined the contribution of individual proteins in the PCs by spectral-counting label-free analysis, a mass spectrometry technique widely used for quantitative analysis of proteins, recently employed by others for PCs of NPs.[48–50] For each patient (AD and healthy), the percentage contribution was calculated 6 times, that is, once for each NP (full raw data are provided in Tables S5 and S6, Supporting Information, Supporting Information). Unlike the practice in classic biomarker discovery experiments, we did not check for single proteins exclusively identified in the PC of AD to discriminate disease samples from healthy ones. The combination of the PCs of the 6 NPs generated an AD-specific PC identifier that was different from that of healthy individuals.

2.3. Machine-Learning Outcomes

The mass spectrometry data from the nanoplatform's protein coronas were then fed into a machine learning algorithm to dis-

criminate AD and early-AD protein coronas from H-PCs. The 19 plasma samples (11 AD samples and 8 healthy samples) were randomly stratified into two sets: A training set of 10 samples (6 AD and 4 healthy) and a test set of 9 samples (5 AD and 4 healthy). The training set was used to train a random forest classifier algorithm[51,52] designed to categorize protein corona mass spectrometry data into either "AD" or "healthy." Low-rank tensor factorization[53–55] (discussed in more depth in Section 4) was utilized to process and de-noise the spectrometry data before classifier training. This de-noising implicitly decreases the considerable variability noticed in individual corona elements. The processed spectrometry data were then used to train the random forest classifier. We then trained a random forest classifier, a popular non-linear classification algorithm,[51,52] on the resulting de-noised data.

We then tested the accuracy of this classifier on the test set, for the purpose of correctly labeling these blind samples as "AD" or "healthy." We measured the sensitivity and specificity of our classifier (**Table 1**); we report these with 95% confidence intervals which, given the relatively small number of plasma samples, were calculated via a Bayesian approach (using the Jeffreys noninformative prior). We also measured the receiver operating characteristic (ROC) curve (**Figure 4**), which captures the range of sensitivity and specificity pairs achievable with our classifier; the specific numbers reported in Table 1 can be thought of as a single point on this curve, shown in red in the figure. Finally, we measured the area under the ROC curve, or area under the curve (AUC), which is a standard measure of accuracy for classification tasks.

To reduce the bias in results that may have been introduced by how the data were split between the training and testing sets, the classifier training and its characterization were re-run with the

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
HEALTHCARE
MATERIALS**

www.advhealthmat.de

a)

	NP Type	Bare NPs	Healthy Plasma	AD Plasma
Zeta Potential (mV)	P	-42.17 ± 0.54	-20.36 ± 1.54	-26.56 ± 2.14
	P-NH2	-27.57 ± 1.63	-29.81 ± 1.55	-30.62 ± 0.74
	P-COOH	-49.38 ± 1.1	-34.87 ± 0.78	-27.02 ± 0.71
	S	-53.72 ± 1.05	-23.22 ± 1.55	-21.23 ± 0.54
	S-NH2	-46.44 ± 0.86	-24.26 ± 0.95	-21.21 ± 0.38
	S-COOH	-56.77 ± 0.14	-28.37 ± 0.87	-25.69 ± 1.31
Size (nm)	P	91.6 ± 1.2	151.9 ± 1.3	113.7 ± 2.5
	P-NH2	100.6 ± 1.7	136.4 ± 3.5	145.3 ± 10.9
	P-COOH	96.2 ± 1.7	117.8 ± 3.1	139.1 ± 10.5
	S	89.4 ± 2.2	136.5 ± 2.2	113.1 ± 4.7
	S-NH2	80.7 ± 1.2	119.7 ± 2.6	105.3 ± 5.8
	S-COOH	99.6 ± 2.4	123.1 ± 1.8	141.7 ± 24.8

b)

Figure 2. Physicochemical properties of NPs before and after incubation with human plasmas. A) Surface charge and size of NPs before and after incubation with AD and control plasmas. Results are reported as average ± SD (*n* = 3). B) TEM analysis. NPs before and after coating with AD protein coronas, analyzed by transmission electron microscopy to evaluate potential changes in morphology and size. All the NPs show a size increase following incubation in plasma.

Table 1. Non-cohort sensitivity, specificity, and area under ROC curve.

Array size	Sensitivity [%]	Specificity [%]	AUC [%]
1	90.3	83.7	92.89
	[5298]	[4295]	
2	98.1	90.4	98.92
	[60 100]	[4897]	
3	98.5	91.3	99.40
	[60 100]	[4997]	
4	99.6	91.6	99.65
	[62 100]	[4997]	
5	99.4	92.8	99.74
	[61 100]	[5098]	
6	100.0	93.7	99.80
	[62 100]	[51.98]	

Classification accuracy for PC nanosystem with array size increasing from 1 to 6 nanoparticles (column 1); Sensitivity and specificity, along with associated confidence intervals, improve with additional NPs (columns 2–3); Area under the ROC curve also increases with array size (column 4); Experimental results are averaged over 1000 independent draws of a training set comprising 10 plasmas, with evaluation on the remaining 9 plasmas; Mean values, along with 95% confidence intervals, are displayed.

training and testing sets randomly chosen in 1000 replications between >10 000 stratified partitions. All reported values are averaged over these 1000 replications, with an average sensitivity of 99.96% (rounded to 100.0% in the table), average specificity of 93.7%, and an average AUC of 99.8%.

The classifier training procedure of 1000 replications with random draws of training versus testing set was then repeated with data from only 1 NP instead of 6 to quantify the increase in robustness introduced by the multi-NP platform. Table 1 depicts the differences in sensitivity, specificity, and AUC in a single-NP platform for each of the 6 NPs. The single-NP platform displayed statistically significantly lower sensitivity and specificity with $p < 10^{-119}$ in a one-sided Wilcoxon signed-rank test in comparison to the 6-NP multi-nanoplatform (Table S2, Supporting Information). The robustness versus number of NPs was further analyzed by repeating the procedure with a nanoplatform consisting of 2 to 5 NPs (Table 1; rows 2–5). For the most part, the nanoplatform improves with more nanoparticles, with the highest sensitivity and specificity achieved by the full 6-NP system (Table S2, Supporting Information; FDR < 1%). This confirms that extracting different PC profiles with multiple NPs with different surface properties drives the robustness of our nanoplatform. Since our aim was also to verify the performance of this platform for

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED
HEALTHCARE
MATERIALS
www.advhealthmat.de
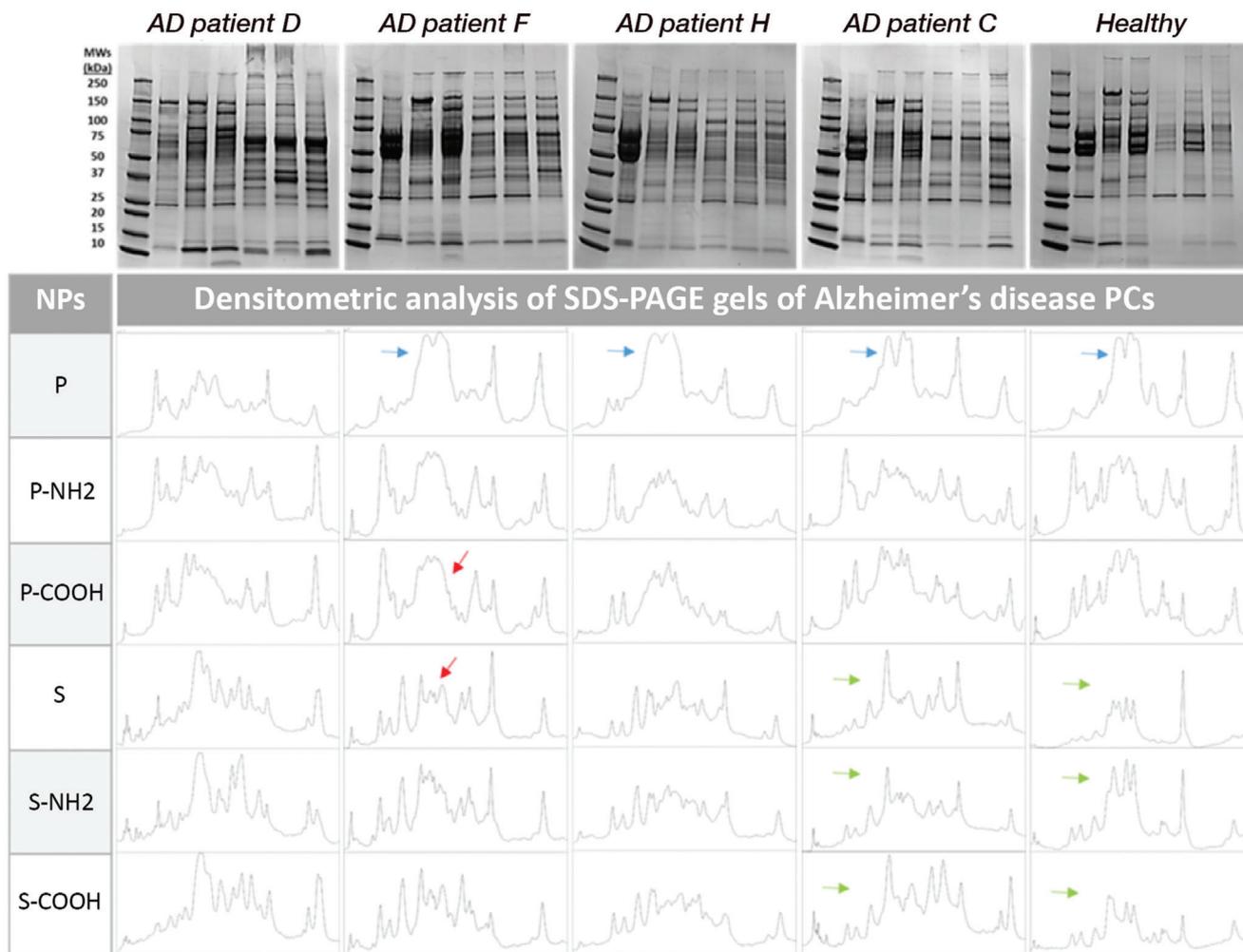
Figure 3. SDS-PAGE gels and densitometric analysis of protein corona profiles. PC profiles analyzed and compared through SDS-PAGE. 4 representative gels of AD PC and 1 healthy PC are shown. Loading order: P, P-NH2, P-COOH, S, S-NH2, S-COOH. Intensity of bands relative to plasma proteins adsorbed on NPs was analyzed by ImageJ (y-axis: intensity, x-axis: molecular weight).
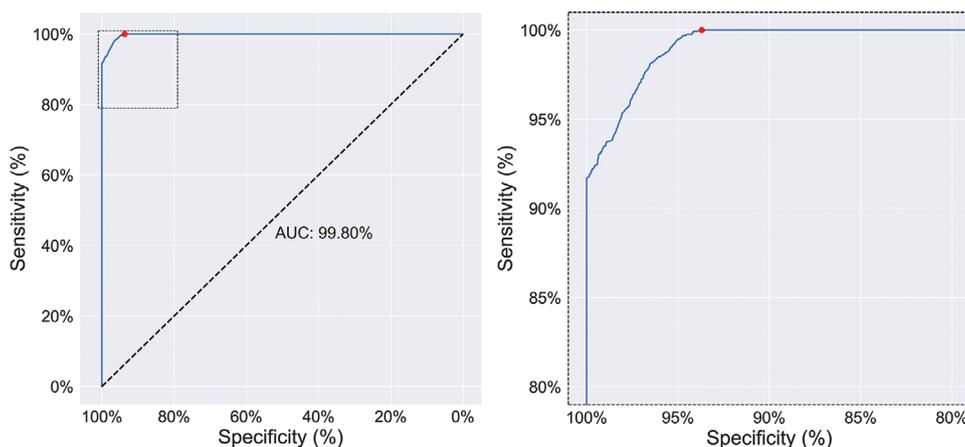
Figure 4. Performance of full nanoparticle array on non-cohort. Left) ROC curve for full PC nanosystem with 6 nanoparticles on non-cohort samples. Area under this curve is equal to 99.80%. The red point indicates the reported sensitivity and specificity, and the rest of the curve represents potential tradeoffs between sensitivity and specificity. Right) The same ROC curve, zoomed into the top left portion. Experimental results are averaged over 1000 independent draws of a training set comprising 10 plasmas, with evaluation on the remaining 9 plasmas.

**ADVANCED
SCIENCE NEWS**
www.advancedsciencenews.com

**ADVANCED
HEALTHCARE
MATERIALS**
www.advhealthmat.de

Table 2. Cohort sensitivity, specificity, and area under ROC curve.

Array size	Sensitivity [%]	Specificity [%]	AUC [%]
1	92.0	71.2	93.43
	[6398]	[2193]	
2	96.5	86.3	98.09
	[6999]	[2797]	
3	96.5	87.4	98.46
	[6999]	[2798]	
4	97.1	88.9	98.85
	[7099]	[2898]	
5	97.5	91.4	99.22
	[70 100]	[2999]	
6	97.6	95.5	99.47
	[7199]	[3199]	

Classification accuracy for PC nanosystem with array size increasing from 1 to 6 nanoparticles (column 1); Sensitivity and specificity, along with associated confidence intervals, improve with additional NPs (columns 2–3); Area under the ROC curve also increases with array size (column 4); Experimental results are averaged over 1000 independent draws of a training set comprising 11 plasmas, with evaluation on the remaining 10 plasmas; Mean values, along with 95% confidence intervals, are displayed.

early diagnosis, we also included in this analysis samples from a cohort study: Plasma of individuals healthy at the time of blood collection who developed AD several years later. We followed the procedure described above, with 1000 replications with random splits between training and testing sets. The splitting in this instance consisted of 11 training samples (8 AD, 3 healthy) and 10 test samples (8 AD, 2 healthy). Raw mass spectrometry data were processed and de-noised as described above, followed by the random forest classifier. Average sensitivity and specificity were 97.6% and 95.5%, respectively (**Table 2**), and the average AUC was 99.47 (**Figure 5**). Given that both values exceeded 95%, we posit that the nanoplatform developed here successfully detects early-onset AD.
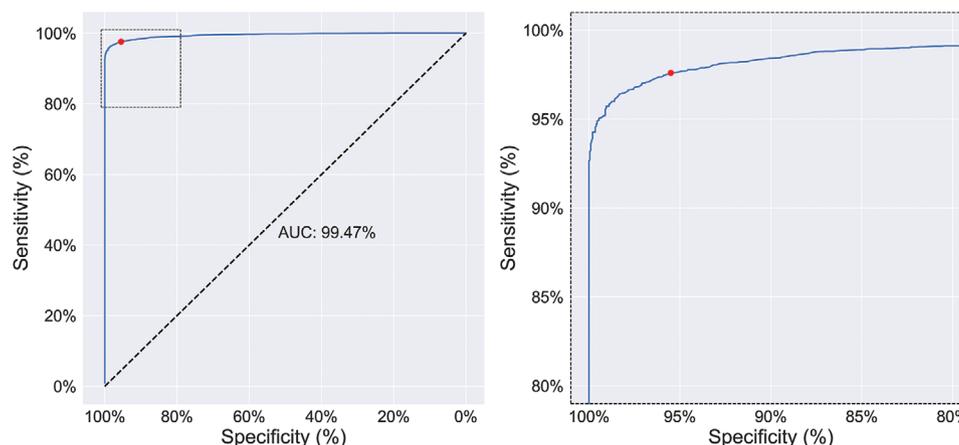
2.4. Bioinformatics Outcomes

The random forest model yields an importance score for each variable (i.e., each NP-protein pair). This score is related to the significance of that variable in discriminating healthy versus AD samples (described in detail in Section 4) and allows us to identify the proteins that are overall most important to our model. We sorted proteins by their model importance score—**Figure 6** and Table S3, Supporting Information, show the 30 proteins with highest model importance score for the non-cohort samples, and **Figure 7** and Table S4, Supporting Information, show the 30 highest-scoring proteins for the cohort samples. These proteins were detected by combining all 6 NPs. To evaluate the robustness of our model, we calculated these importance scores across the previously described 1000 random splits of the data, to determine whether the relative importance of these proteins is stable. Figures 6 and 7 also show, for each importance score, the 25th to 75th percentiles of that score across the 1000 splits of the data. The relatively small size of these intervals demonstrates that the set of important proteins is robust to the split of data used for model training, illustrating model stability.

Furthermore, to gain insight into the biological explanation behind the machine learning-facilitated discovery of proteins significant in AD, we interrogated the Open Targets database (https://www.opentargets.org/; date accessed: 28 June 2018). Open Targets integrates public databases such as GWAS Catalog, UniProt, Gene2Phenotype, Cancer Gene Census, IntOGen, Europe PMC, and Reactome to calculate an association between a protein and a disease on a scale of 0.0 (weakest) to 1.0 (strongest).[56] Using the Experimental Factor Ontology (EFO) disease terms of "neurodegenerative disease (EFO:0 005772)" and "Alzheimers disease (EFO:0000249)" (a child in the ontology tree of neurodegenerative diseases), we queried Open Targets for all the associated targets and respective scores. Of the 444 total proteins identified in the non-cohort sample, 8 have an Open Targets score denoting a strong association (above 0.8) with neurodegenerative disease. By comparison, of the 681 total proteins identified in the cohort sample, 20 have a strong association with neurodegenerative

Figure 5. Performance of full nanoparticle array on cohort. Left) ROC curve for full PC nanosystem with 6 nanoparticles on cohort samples. Area under this curve is equal to 99.47%. The red point indicates the reported sensitivity and specificity, and the rest of the curve represents potential tradeoffs between sensitivity and specificity. Right) The same ROC curve zoomed into the top left portion. Experimental results are averaged over 1000 independent draws of a training set comprising 11 plasmas, with evaluation on the remaining 10 plasmas.

Figure 6. Protein importance and model stability for classification (non-cohort). Random forest importance scores for the top 30 proteins with highest total score (summed across 6 NPs) in the non-cohort classification model. Each column indicates the total classification importance of a single protein, with the 6 colored stacks within a column indicating the separate importances of the observed interaction of that protein with each of the 6 NPs. Each colored stack is accompanied by vertical error bars that indicate the 25th and 75th quantiles of the same importance score across classifiers trained on 1000 random draws of the training set from the data (i.e., the same train-test draws previously used to evaluate the model). The relatively small sizes of these confidence intervals indicate the "stability" of the trained model in terms of the protein-NP interactions upon which it crucially relies, with respect to random draws of data. The proteins are listed in Table S3, Supporting Information.
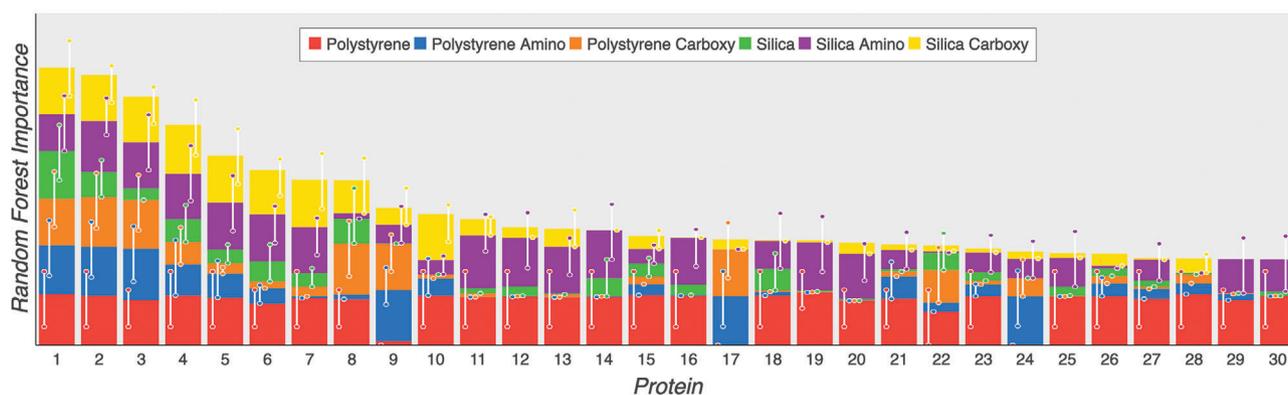
Figure 7. Protein importance and model stability for classification (cohort). Random forest importance scores for the top 30 proteins with highest total score (summed across 6 NPs) in the cohort classification model. Each column indicates the total classification importance of a single protein, with the 6 colored stacks within a column indicating the separate importances of the observed interaction of that protein with each of the 6 NPs. Each colored stack is accompanied by vertical error bars that indicate the 25th and 75th quantiles of the same importance score across classifiers trained on 1000 random draws of the training set from the data (i.e., the same train-test draws previously used to evaluate the model). The relatively small sizes of these confidence intervals indicate the "stability" of the trained model in terms of the protein-NP interactions upon which it crucially relies, with respect to random draws of data. The proteins are listed in Table S4, Supporting Information.

disease. Among these, Apolipoprotein E (P02649/APOE) is very well known to be associated with late-onset sporadic AD, particularly its E4 variant, which represents the most well-known genetic risk factor in many ethnicities.[57] This protein was present in both sets. In addition, we see P14136/GFAP Glial fibrillary acidic protein, whose gene mutations cause Alexander disease, a rare disorder of astrocytes in the central nervous system. P05067/APP Amyloid beta precursor protein, a well-known target for Alzheimer's disease, was present only in the cohort sample.[57]

Though the high Open Targets scoring proteins vary in their Random Forest classification importance, the sample size is too small to calculate accurate absolute or relative model performance.

3. Conclusion

Here we demonstrate that PCs from six different NPs could be used to train a classifier machine-learning algorithm to diagnose both AD and early-onset AD with significant accuracy and specificity. While CSF-based and single-biomarkers blood-based tests suffer from high rates of false positives and require further complementary analyses to reach a final diagnosis, this approach can record the interactions between NPs and all the specific high-affinity proteins in a biological fluid, for example, plasma, allowing the generation of a highly specific PC identifier for AD.

In the literature, the combination of CSF Aβ1-42 and T-tau at baseline yielded sensitivity and specificity levels of 95% and 83%, respectively, for clinical AD diagnosis in patients with MCI.[58]

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
HEALTHCARE
MATERIALS**

www.advhealthmat.de

On the other hand, the PET imaging system was able to discriminate between Alzheimer's and other dementias with > 90% sensitivity and 70% specificity.[59] Our nano-array system's sensitivity and specificity for AD detection from healthy cohorts were ≈100% and > 93%, respectively.

To reiterate, our approach is not based on the identification of specific, predetermined biomarkers. The NPs concentrate only those proteins showing the highest affinity toward their surface, thus allowing the detection of proteins that undergo only slight changes in abundance in the presence of a disease. In other words, we can say that our library of NPs interacts with hundreds of proteins in the interrogated biofluid in an unbiased way, meaning that this approach is not a targeted or an antibody-based method. The compositions of the PCs around the NPs are used to discriminate between samples.

Indeed, the nanoarray system can be easily modified to increase robustness and predictive value by using additional nanoparticles with various chemical properties. In addition, while reliable detection of low-abundance proteins is difficult with assays like ELISA, the nanoarray system allows such low-abundance proteins not only to be detected due to the high surface area and adsorbing nature of nanoparticles, but also to play a role in the diagnostic process if the random forest algorithm deems them an important feature for classification.

We believe that our findings substantially advance the field of neurodegenerative nanotechnology and may pave the way for early diagnostic approaches, including the detection of devastating diseases such as AD, long before they can be diagnosed through currently available screening approaches or symptoms. Different from the current standard biomarker tests, whose procedure for extraction is very invasive and discourages asymptomatic patients from getting early and routine testing, our approach is based only on blood serum. It eliminates a major roadblock to routine testing and early detection, akin to the liquid biopsies for cancer that were recently approved by the FDA. To further improve the test's performance, we plan to focus on selecting an ideal set of different NPs, both number-wise and type-wise, that identifies the highest number of proteins and offers the best performance in terms of accuracy and ease of execution. Moreover, the nanoplatform could serve as an alternative to current plasma proteome-based studies in the discovery of novel, low-concentration protein biomarkers.

4. Experimental Section

Nanoparticles: 90–100 nm plain, amino-conjugated and carboxyl-conjugated silica nanoparticles were supplied by Kisker-Products (https://www.kisker-biotech.com/). 90–100nm plain, amino-conjugated, and carboxyl-conjugated polystyrene nanoparticles were supplied by Polysciences, Inc. (http://www.polysciences.com/). The nanoparticles' size distribution, morphology, and surface charge (zeta potential) were characterized as described in the following sections.

Personalized Protein Corona Formation: Human plasma from Alzheimer's disease patients and healthy individuals was purchased from Innovative Research, Inc. (Novi, MI, USA). To create protein coronas around NPs, plasmas from 11 current AD patients, eight healthy individuals, and 16 AD patients from a cohort study were used (healthy at the time of withdrawal, years later diagnosed with AD) with five controls. The cohort plasma samples were collected through the NIH-funded Golestan Cohort Study, performed by the National Cancer Institute (NCI) in the

USA, the International Agency for Research on Cancer (IARC) in France, and the Tehran University of Medical Sciences (TUMS) in Iran. This study involved the collection and storage of plasma from 50 000 healthy subjects, over 1000 of whom went on to develop various diseases.

Protein coronas were created by incubating NPs in a solution with a 1:1 volume ratio of deionized H_2O and human plasma at 37 °C for 1 h under constant agitation. Immediately after incubation, PC-coated NPs were centrifuged (14 000rpm and 10 °C for 30 min) and subjected to extensive washing in cold phosphate-buffered saline to remove soft PC. PC-coated NPs were resuspended in denaturing buffer for SDS-PAGE gels and LC/MSMS analysis, or in deionized H_2O for size distribution and surface charge characterization.

Physicochemical Characterization of Nanoparticles: Size distribution was characterized with nanoparticle tracking analysis (Nanosight, Malvern, UK), which calculates the sizes of nanoparticles flowing through a canal with a scattered laser beam detected on an optical microscope. Meanwhile, a video was recorded by a camera aligned to the beam showing the movement of the NPs (30-60 frame/sec). ζ-potential of bare and protein corona-coated NPs was determined using a Zetasizer Nano ZS90 (Malvern, UK). NPs were diluted in bidistilled water before the analysis to a concentration of 50 µg mL^{-1}. Size and surface charge values are given as mean ± SD of three independent measurements. For TEM analysis, samples and grids were labeled with 1% uranyl acetate. A Tecnai G2 Spirit BioTWIN Transmission Electron Microscope equipped with an AMT 2k CCD camera was used.

1D Gel Electrophoresis: PC was dissolved in 8 M urea and 50 mM ammonium bicarbonate. An equal volume of Laemmli buffer 2X was added and boiled for 5 min at 90 °C. Samples were resolved onto 4–20% Mini-PROTEAN TGX Precast Gels (Bio-Rad Laboratories, Hercules, CA) for 1 h at 120 V. Coomassie Brilliant Blue (Fisher Scientific, Fair Lawn, NJ, USA) overnight staining followed by washing in ultra-pure water was carried out for protein visualization. Densitometric analysis of protein bands was performed using ImageJ (website: imagej.nih.gov/ij/).

Protein Identification and Quantification by Mass Spectrometry: Proteins were treated with 10 mM dithiothreitol (Sigma) for 1 h at 56 °C and 55 mM iodoacetamide (Sigma-Aldrich, St Loius, MO, USA) for 1 h at 25 °C in the dark. Proteins were then hydrolyzed with trypsin (Promega, Madison, WI, USA) at an enzyme/substrate ratio of 1:50 in 100 mM ammonium acetate, pH 8.9 at 25 °C overnight. A 5% acetic acid (99.9%, Sigma-Aldrich) solution was utilized to deactivate the trypsin. Peptides were desalted using C18 SpinTips (Protea, Morgantown, WV), then vacuum centrifuged and stored at −80 °C until the day of analysis. Peptides were separated by reverse-phase HPLC (Thermo Fisher, Waltham, MA Easy nLC1000) using a precolumn (made in house, 6 cm of 10 µm C18) and a self-pack 5 µm tip analytical column (12 cm of 5 µm C18, New Objective) over a 140-min gradient before nano-electrospray using a QExactive mass spectrometer (Thermo Fisher). Solvent A was 0.1% formic acid and solvent B was 80% MeCN/0.1% formic acid. The gradient conditions were 2–10% B (0–3 min), 10–30% B (3–107 min), 30–40% B (107–121 min), 40–60% B (121–126 min), 60–100% B (126–127 min), 100% B (127–137 min), 100–0% B (137–138 min), and 0% B (138–140 min), and the mass spectrometer was operated in data-dependent mode. The parameters for the full-scan MS were: Resolution of 70 000 across 350–2000 m/z, AGC 3e6, and maximum IT 50 ms. The full MS scan was followed by MS/MS for the top 10 precursor ions in each cycle with a normalized collision energy of 28 and dynamic exclusion of 30 s. Raw mass spectral data files (.raw) were searched using Proteome Discoverer (Thermo Fisher) and Mascot version 2.4.1 (Matrix Science). Mascot search parameters were: 10 ppm mass tolerance for precursor ions; 15 millimass units (mmu) for fragment ion mass tolerance; 2 missed cleavages of trypsin; fixed modification was carbamidomethylation of cysteine; the only variable modification was methionine oxidation. Only peptides with a Mascot score ≥ 25 were included in the data analysis. Spectral counting was performed by summing the total number of peptides selected for fragmentation of each protein.

Statistical Analysis: All statistical analyses were performed in Python using the scikit-learn, numpy, and scipy packages, and figures and graphs were created using the bokeh package in Python, along with Microsoft Excel, XLSTAT, and MATLAB. For all plasma samples, a data matrix X_i was

generated such that each row of the matrix corresponded to the protein abundances of a single nanoparticle, as obtained from the protein corona nanoplatform. As a preprocessing step, the protein abundances were converted to relative protein abundances by normalizing the rows of all of the matrices.

Tensor Factorization: The data were treated as a three-mode tensor, the first two modes corresponding to nanoparticles and proteins, and the third mode corresponding to plasma samples; this was essentially equivalent to stacking the observation matrices corresponding to each sample, X_i, on top of each other. The data were de-noised via a low Tucker rank tensor factorization[44–46] using code implemented in Python for this project (available for academic use upon request). Each matrix X_i, was approximated by a tensor decomposition that takes the form

$$X_i \approx US_iV^{\mathsf{T}} \tag{1}$$

where U is a matrix whose rows can be viewed as latent features corresponding to each of the nanoparticles, and similarly V is a matrix whose rows can be viewed as latent features corresponding to each of the proteins; these latent features are shared across all of the data matrices. Finally, each S_i is a matrix encoding interactions between nanoparticle and protein features, and these are allowed to be unique between samples. This decomposition was estimated in two steps: a) U and V were estimated via a truncated singular-value decomposition on the mode-1 and mode-2 unfoldings of the tensor, and then given these estimates, b) each S_i matrix was fit separately via a least-squares calculation.

Random Forest Classification: The random forest model is a well-known machine-learning algorithm for classification. A random forest is made up of multiple decision trees that each make simple classification decisions based on relatively few variables. These trees were created (or "trained") with different, randomly drawn subsets of variables, reducing the likelihood of two identical trees. Given a new sample, each tree was traversed top-down until a set of training samples reached at the bottom. Using the forest as a whole for classification amounts to having the multiple decision trees "vote" on a label (in this case, AD or healthy), where each tree's vote was made from the labels of the bottom set of training samples. For the authors#x00027; own algorithm, each random forest consisted of 1000 decision trees and was trained using the scikit-learn package. Importance scores were also calculated using the same package. On an individual "tree" of the random forest, the importance score of any variable used in constructing the tree was defined as the proportion of the training set that lies in the "leaves" of nodes utilizing that variable (variables not used in constructing the tree are assigned a score of zero); then the overall importance score for a variable is the average of its importance scores on each tree.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Conflict of Interest

O.C.F. and R.L. declare financial interests in Selecta Biosciences, Tarveda Therapeutics, and Seer. V.F., T.P., and P.M. declare financial interests in Seer.

Author Contributions

C.C. designed and conducted the experiments assisted by G.Y.L., S.S., R.M., and S.B.; A.A.L. and V.F. provided machine learning analysis; H.P. provided guidance on the cohort data; C.C. wrote the paper and revised it according to the comments of authors; T.P. and P.M. provided technical support; R.L. provided conceptual advice; O.C.F. directed the project and provided mentoring and funding support.

[1] M. Prince, R. Bryce, C. Ferri, *World Alzheimer Report 2011: The Benefits of Early Diagnosis and Intervention: Alzheimer's Disease International*, London, **2011**.
[2] Alzheimer's Association, *Alzheimer's Dementia* **2019**, *15*, 321.
[3] J. A. Hendrix, R. J. Bateman, H. R. Brashear, C. Duggan, M. C. Carrillo, L. J. Bain, R. DeMattos, R. G. Katz, S. Ostrowitzki, E. Siemers, R. Sperling, O. V. Vitolo, *Alzheimer's Dementia* **2016**, *12*, 623.
[4] P. Schneider, H. Hampel, K. Buerger, *CNS Neurosci. Ther.* **2009**, *15*, 358.
[5] C. R. Jack, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue, J. Q. Trojanowski, *Lancet Neurol.* **2013**, *12*, 207.
[6] E. Mcdade, R. J. Bateman, *Nature* **2017**, *547*, 153.
[7] V. L. Villemagne, S. Burnham, P. Bourgeat, B. Brown, K. A. Ellis, O. Salvado, C. Szoeke, S. L. Macaulay, R. Martins, P. Maruff, D. Ames, C. C. Rowe, C. L. Masters, *Lancet Neurol.* **2013**, *12*, 357.
[8] T. L. S. Benzinger, T. Blazey, C. R. Jack, R. A. Koeppe, Y. Su, C. Xiong, M. E. Raichle, A. Z. Snyder, B. M. Ances, R. J. Bateman, N. J. Cairns, A. M. Fagan, A. Goate, D. S. Marcus, P. S. Aisen, J. J. Christensen, L. Ercole, R. C. Hornbeck, A. M. Farrar, P. Aldea, M. S. Jasielec, C. J. Owen, X. Xie, R. Mayeux, A. Brickman, E. Mcdade, W. Klunk, C. A. Mathis, J. Ringman, P. M. Thompson, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E4502.
[9] A. M. Fagan, C. Xiong, M. S. Jasielec, R. J. Bateman, A. M. Goate, T. L. S. Benzinger, B. Ghetti, R. N. Martins, C. L. Masters, R. Mayeux, J. M. Ringman, M. N. Rossor, S. Salloway, P. R. Schofield, R. A. Sperling, D. Marcus, N. J. Cairns, V. D. Buckles, J. H. Ladenson, J. C. Morris, D. M. Holtzman, *Sci. Transl. Med.* **2014**, *6*, 226ra30..
[10] B. Dubois, H. H. Feldman, C. Jacova, H. Hampel, J. L. Molinuevo, K. Blennow, S. T. Dekosky, S. Gauthier, D. Selkoe, R. Bateman, S. Cappa, S. Crutch, S. Engelborghs, G. B. Frisoni, N. C. Fox, D. Galasko, M.-O. Habert, G. A. Jicha, A. Nordberg, F. Pasquier, G. Rabinovici, P. Robert, C. Rowe, S. Salloway, M. Sarazin, S. Epelbaum, L. C. De Souza, B. Vellas, P. J. Visser, L. Schneider, Y. Stern, P. Scheltens, J. L. Cummings, *Lancet Neurol.* **2014**, *13*, 614.
[11] H. Zetterberg, *J. Neurosci. Methods* **2019**, *319*, 2..
[12] H. Zetterberg, J. M. Schott, *Nat. Med.* **2019**, *25*, 201.
[13] K. Henriksen, S. E. O'bryant, H. Hampel, J. Q. Trojanowski, T. J. Montine, A. Jeromin, K. Blennow, A. Lönneborg, T. Wyss-Coray, H. Soares, C. Bazenet, M. Sjögren, W. Hu, S. Lovestone, M. A. Karsdal, M. W. Weiner, *Alzheimer's Dementia* **2014**, *10*, 115.
[14] M. J. Hajipour, M. R. Santoso, F. Rezaee, H. Aghaverdi, M. Mahmoudi, G. Perry, *Trends Biotechnol.* **2017**, *35*, 937.
[15] I. Lynch, K. A. Dawson, *Nano Today* **2008**, *3*, 40.
[16] T. Cedervall, I. Lynch, S. Lindman, T. Berggard, E. Thulin, H. Nilsson, K. A. Dawson, S. Linse, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2050.
[17] I. Lynch, A. Salvati, K. A. Dawson, *Nat. Nanotechnol.* **2009**, *4*, 546.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
HEALTHCARE
MATERIALS**

www.advhealthmat.de

[18] M. Rahman, S. Laurent, N. Tawil, L. H. Yahia, M. Mahmoudi, *Protein-Nanoparticle Interactions*, Springer, Berlin **2013**, pp. 21–44.

[19] C. D. Walkey, W. C. W. Chan, *Chem. Soc. Rev.* **2012**, *41*, 2780.

[20] M. P. Monopoli, C. Åberg, A. Salvati, K. A. Dawson, *Nat. Nanotechnol.* **2012**, *7*, 779..

[21] C. Corbo, R. Molinaro, A. Parodi, N. E. Toledano Furman, F. Salvatore, E. Tasciotti, *Nanomedicine* **2016**, *11*, 81..

[22] C. Corbo, R. Molinaro, F. Taraballi, N. E. T. Furman, M. B. Sherman, A. Parodi, F. Salvatore, E. Tasciotti, *Int. J. Nanomed.* **2016**, *11*, 3049.

[23] A. Lesniak, F. Fenaroli, M. P. Monopoli, C. Åberg, K. A. Dawson, A. Salvati, *ACS Nano* **2012**, *6*, 5845.

[24] S. Tenzer, D. Docter, J. Kuharev, A. Musyanovych, V. Fetz, R. Hecht, F. Schlenk, D. Fischer, K. Kiouptsi, C. Reinhardt, K. Landfester, H. Schild, M. Maskos, S. K. Knauer, R. H. Stauber, *Nat. Nanotechnol.* **2013**, *8*, 772.

[25] M. P. Monopoli, D. Walczyk, A. Campbell, G. Elia, I. Lynch, F. B. Bombelli, K. A. Dawson, *J. Am. Chem. Soc.* **2011**, *133*, 2525.

[26] M. J. Hajipour, O. Akhavan, A. Meidanchi, S. Laurent, M. Mahmoudi, *RSC Adv.* **2014**, *4*, 62557.

[27] G. Caracciolo, O. C. Farokhzad, M. Mahmoudi, *Trends Biotechnol.* **2017**, *35*, 257.

[28] M. Safi, J. Courtois, M. Seigneuret, H. Conjeaud, J.-F. Berret, *Biomaterials* **2011**, *32*, 9353.

[29] C. Corbo, R. Molinaro, F. Taraballi, N. E. Toledano Furman, K. A. Hartman, M. B. Sherman, E. De Rosa, D. K. Kirui, F. Salvatore, E. Tasciotti, *ACS Nano* **2017**, *11*, 3262.

[30] M. R. Sepand, M. Ghavami, S. Zanganeh, S. Stacks, F. Ghasemi, H. Montazeri, C. Corbo, H. Derakhshankhah, S. N. Ostad, M. H. Ghahremani, M. Mahmoudi, *Nanoscale* **2020**, *12*, 4935.

[31] F. Hashemi, M. R. Hormozi-Nezhad, C. Corbo, F. Farvadi, M. A. Shokrgozar, M. Mehrjoo, F. Atyabi, M. H. Ghahremani, M. Mahmoudi, R. Dinarvand, *Nanoscale* **2019**, *11*, 5974.

[32] M. J. Hajipour, S. Laurent, A. Aghaie, F. Rezaee, M. Mahmoudi, *Biomater. Sci.* **2014**, *2*, 1210..

[33] V. Colapicchioni, M. Tilio, L. Digiacomo, V. Gambini, S. Palchetti, C. Marchini, D. Pozzi, S Occhipinti, A. Amici, G. Caracciolo, *Int. J. Biochem. Cell Biol.* **2016**, *75*, 180.

[34] C. Corbo, R. Molinaro, M. Tabatabaei, O. C. Farokhzad, M. Mahmoudi, *Biomater. Sci.* **2017**, *5*, 378.

[35] T. Zheng, N. Pierre-Pierre, X. Yan, Q. Huo, A. J. Almodovar, F. Valerio, I. Rivera-Ramirez, E. Griffith, D. D. Decker, S. Chen, N. Zhu, *ACS Appl. Mater. Interfaces* **2015**, *7*, 6819.

[36] S. Xu, T. Gao, X. Feng, X. Fan, G. Liu, Y. Mao, X. Yu, J. Lin, X. Luo, *Biosens. Bioelectron.* **2017**, *97*, 203.

[37] M. Rahman, M. Mahmoudi, *Disease Specific Protein Corona*, SPIE, Bellingham, USA **2015**.

[38] T. Zheng, N. Pierre-Pierre, X. Yan, Q. Huo, A. J. O. Almodovar, F. Valerio, I. Rivera-Ramirez, E. Griffith, D. D. Decker, S. Chen, N. Zhu, *ACS Appl. Mater. Interfaces* **2015**, *7*, 6819.

[39] T. Zheng, C. Finn, C. J. Parrett, K. Dhume, J. H. Hwang, D. Sidhom, T. M. Strutt, Y. Y. Li Sip, K. K. Mckinstry, Q. Huo, *ACS Infect. Dis.* **2017**, *3*, 866.

[40] D. Caputo, M. Papi, R. Coppola, S. Palchetti, L. Digiacomo, G. Caracciolo, D. Pozzi, *Nanoscale* **2017**, *9*, 349.

[41] J. D. Lapek, R. H. Fang, X. Wei, P. Li, B. Wang, L. Zhang, D. J. Gonzalez, *ACS Nano* **2017**, *11*, 11831.

[42] U. Distler, S. Tenzer, *ACS Nano* **2017**, *11*, 11768.

[43] L. Papafilippou, A. Claxton, P. Dark, K. Kostarelos, M. Hadjidemetriou, *Nanoscale* **2020**, *12*, 10240.

[44] L. Digiacomo, K. Jafari-Khouzani, S. Palchetti, D. Pozzi, A. L. Capriotti, A. Laganà, R. Z. Chiozzi, D. Caputo, C. Cascone, R. Coppola, G. Flammia, V. Altomare, A. Grasso, M. Mahmoudi, G. Caracciolo, *Nanoscale* **2020**, *12*, 16697.

[45] J. E. Blume, W. C. Manning, G. Troiano, D. Hornburg, M. Figa, L. Hesterberg, T. L. Platt, X. Zhao, R. A. Cuaresma, P. A. Everley, M. Ko, H. Liou, M. Mahoney, S. Ferdosi, E. M. Elgierari, C. Stolarczyk, B. Tangeysh, H. Xia, R. Benz, A. Siddiqui, S. A. Carr, P. Ma, R. Langer, V. Farias, O. C. Farokhzad, *Nat. Commun.* **2020**, *11*, 3662..

[46] S. Palchetti, V. Colapicchioni, L. Digiacomo, G. Caracciolo, D. Pozzi, A. L. Capriotti, G. La Barbera, A. Laganà, *Biochim. Biophys. Acta* **2016**, *1858*, 189.

[47] M. Kokkinopoulou, J. Simon, K. Landfester, V. Mailänder, I. Lieberwirth, *Nanoscale* **2017**, *9*, 8858.

[48] D. Docter, U. Distler, W. Storck, J. Kuharev, D. Wünsch, A. Hahlbrock, S. K. Knauer, S. Tenzer, R. H. Stauber, *Nat. Protoc.* **2014**, *9*, 2030.

[49] V. Mirshafiee, M. Mahmoudi, K. Lou, J. Cheng, M. L. Kraft, *Chem. Commun.* **2013**, *49*, 2557.

[50] N. P. Mortensen, G. B. Hurst, W. Wang, C. M. Foster, P. D. Nallathamby, S. T. Retterer, *Nanoscale* **2013**, *5*, 6372..

[51] L. Breiman, *Mach. Learn.* **2001**, *45*, 5.

[52] A. Liaw, M. Wiener, *R News* **2002**, *2*, 18.

[53] T. G. Kolda, B. W. Bader, *SIAM Rev.* **2009**, *51*, 455.

[54] L. De Lathauwer, B. De Moor, J. Vandewalle, *SIAM J. Matrix Anal. Appl.* **2000**, *21*, 1253.

[55] V. Farias, A. Li, in Proc. of the 20th Int. Conf. on Artificial Intelligence and Statistics (Eds: S. Aarti, Z. Jerry), Proceedings of Machine Learning Research, *54*, **2017**, pp. 1394–1402.

[56] D. Carvalho-Silva, A. Pierleoni, M. Pignatelli, C. Ong, L. Fumis, N. Karamanis, M. Carmona, A. Faulconbridge, A. Hercules, E. Mcauley, A. Miranda, G. Peat, M. Spitzer, J. Barrett, D. G. Hulcoop, E. Papa, G. Koscielny, I. Dunham, *Nucleic Acids Res.* **2018**, *47*, D1056.

[57] S. Sadigh-Eteghad, M. Talebi, M. Farhoudi, *Neurosciences* **2012**, *17*, 321..

[58] E. Niemantsverdriet, S. Valckx, M. Bjerke, S. Engelborghs, *Acta Neurol. Belg.* **2017**, *117*, 591..

[59] V. Berti, A. Pupi, L. Mosconi, *Ann. N. Y. Acad. Sci.* **2011**, *1228*, 81..