AI-assisted analysis of flow cytometry data from human hematopoietic cells

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Cathrin Isabel Nollmann

aus Düsseldorf

Düsseldorf, Juli 2024

aus dem Institut für Experimentelle Physik der kondensierten Materie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Thomas Heinzel

2. Prof. Dr. Rainer Haas

Tag der mündlichen Prüfung: 26.08.2024

Abstract

This thesis deals with the investigation of AI-assisted analysis of flow cytometry data from human hematopoietic cells and how this can contribute to clinical diagnostics. In addition, investigations into the establishment of a drug delivery system with carbon nanodots (CNDs) synthesized in the working group were continued. Four manuscripts emerged from this work, which summarize the main findings.

In *Paper I*, the immunological phenotype of hematopoietic stem and progenitor cells (HSPCs) was analyzed using selected markers in comparison to blast and leukemic stem cells (LSCs) from patients with acute myeloid leukemia (AML) and myelodysplastic neoplasms (MDS) assisted by the t-SNE algorithm. The gates defined on the t-SNE map could be assigned to diagnostically relevant cell populations and revealed a fine substructure within the respective cell types. For the selected markers it has been shown that the differences between HSPCs in comparison to blasts, and LSCs in AML and MDS are likely to be due to the different proportions of certain cell types and different intensities of fluorescent markers within the various cell types, rather than the exclusive presence of certain cell types. In addition, a method was introduced for classifying new samples using t-SNE reference images and a quantitative similarity comparison with the pearson coefficient as a measure. This concept can also be used to monitor the evolution of cell populations of patients in therapy and is not limited to the diagnosis of AML and MDS, but can be applied to multiparameter diagnostic flow cytometry data as well as single-cell data in general.

As a further parameter in the AI-supported evaluation of hematopoietic flow cytometry data, the uptake of CNDs was investigated in *Paper II* in comparison between healthy donors and patients with AML. The differential uptake of CNDs between different cell types as well as between the two comparison groups could be shown concisely using t-SNE. While all cell types took up the CNDs, the $CD34^+$ and $CD33^+$ subsets of the AML samples showed a significantly reduced uptake. Confocal fluorescence microscopy images showed that the CNDs accumulate perinuclearly in the AML cell line HL-60, indicating localization in the lysosomes, in agreement with previous studies.

In *Paper III*, the possibilities of functionalizing the CNDs for selective uptake were investigated. For this purpose, monosaccharides and glycooligomers were covalently coupled to the CNDs and the uptake was investigated in multiple cell lines. No cell type-dependent uptake was observed, but the CNDs conjugated with sugar monomers showed a two- to threefold increase in uptake compared to the pristine CNDs and the CNDs functionalized with glycooligomers.

Finally, *Paper IV* shows the results of the study on the influence of CNDs on the lysosomes and the associated process of autophagy, as well as the targeted delivery of drugs into the lysosomes using CNDs as carriers. Lysosomal processes and autophagy as characterized by cathepsin B and L and the autophagy markers SQSTM1/p62 and LC3 were not significantly altered due to the presence of the CNDs, making them promising inert carriers for lysosomal drug delivery. Branched Polyethylenimine (bPEI) was coupled to the CNDs as an example of a drug. bPEI was successfully delivered into lysosomes by the CNDs, confocal fluorescence microscopy revealing increased accumulation of bPEI-CNDs in lysosomes compared to pristine CNDs. In addition, it was observed that the effects of free bPEI on the cell were attenuated by binding to the CNDs.

Eidesstattliche Erklärung

Ich versichere an Eides statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist.

_____   _____
Ort, Datum   Cathrin Isabel Nollmann

Contents

Introduction

Medical research and practice are facing a profound transformation in the age of big data [1]. Especially in the area of single cell data as they may occur in the fields of genomics, proteomics, transcriptomics, and cytomics the amount and complexity of data is increasing due to technological progress [2], [3]. This variety and wealth of information holds tremendous potential for making diagnoses more precise, developing personalized therapies and increasing the efficiency of the healthcare system [4], [5]. However, the sheer volume and complexity of such medical data exceeds the capacity of conventional analysis and processing methods [6]. This creates an urgent need for advanced, AI-supported analysis methods to extract the relevant information from the data and transform it into clinically relevant insights [6], [7].

Flow cytometry is a method of analysis at single cell level [8]. It is used as standard in medical diagnostics for hematological diseases and allows the analysis of up to 10.000 cells per second [9]–[15]. While the first flow cytometer could only record one parameter, namely the size of cells, modern flow cytometers are able to measure up to 30 fluorescence channels simultaneously [16], [17]. The conventional data evaluation is carried out manually by a human operator [18]. Two parameters are plotted against each other and regions of interest, so-called gates, are drawn in [19]. The cell subsets from within the gates can then be further analyzed using other parameter combinations. However, due to the increasing number of parameters that can be measured simultaneously, manual evaluation reaches its practical limits [19]. With 18 parameters, a total of 153 two-dimensional plots are required to display all marker combinations [18]. Manual analysis thus has several drawbacks [17]. Due to the increasing number of markers, the analysis gets even more complex and time-consuming, and when analyzing subspaces defined by two markers, high-dimensional patterns can get lost during the analysis [17], [18]. Furthermore, manual analysis depends on the operator's experience, which may add bias to the analysis so that new information or unknown cell populations may not be identified [18], [20]. AI based tools seem promising to help address these challenges [17], [19], [20]. One method that has already been applied successfully many times to visualize and evaluate multiparameter single cell data sets is based on the t-distributed stochastic neighbor embedding (t-SNE) algorithm [17], [18], [21]–[23].

While analysis and diagnostics continue to develop, therapeutic approaches are also evolving towards ever more precise treatment approaches. One promising approach is using nanoparticles to build a drug delivery system in which the drugs are targeted to their site of action and unwanted side effects can be minimized thanks to the high selectivity [24], [25]. Fluorescent carbon nanodots have proven to be promising candidates for biomedical applications and have already been successfully used for intracellular sensing [26]–[29], drug delivery [30], [31], mRNA delivery [32], and photodynamic therapy [33]–[36].

The aim of this work is to investigate how the use of AI algorithms such as t-SNE can con-

tribute to the analysis of multiparameter flow cytometry data and how it can be used to support clinical diagnostics. Thereby we focused on the data sets of human hematopoietic cells. As an additional parameter in the investigation, we examined the uptake of carbon nanoparticles to differentiate between healthy and diseased samples. The study resulted in two further studies on the uptake of carbon nanodots. As the previous investigations showed that there were differences in the uptake of the particles, an attempt was made on the one hand to make the uptake more selective by modifying the particles. On the other hand, it was shown that the particles accumulated mainly in the lysosomes, which raised the question of whether the particles have an influence on the lysosomes and the associated process of autophagy and whether the particles can be used as a carrier for drug transport into the lysosomes.

This work is divided into five chapters. Chapter 1 sets out the theoretical basis for the work. First, artificial intelligence and machine learning are discussed and applications of dimensionality reduction and clustering are described in more detail. Various applications in the field of medicine are presented, including a separate discussion of applications in the field of flow cytometry. Next, carbon nanoparticles are presented, with specific reference to graphene, graphene quantum dots, carbon nanodots and their biomedical applications. A discussion of the biomedical basics, namely hematopoiesis, CD antigens, the hematologic diseases acute myeloid leukemia (AML) and myelodysplastic neoplasms (MDS), the endolysosomal system and the process of autophagy concludes chapter 1. In chapter 2, the most important experimental methods are briefly explained, while the the specific methodological details are referred to in the respective papers. Starting the chapter, flow cytometry is introduced and the process of gating is discussed. The synthesis and modification of the carbon nanodots is explained next. The following part covers the explanation of confocal fluorescence microscopy, western blot and mass spectrometry. Chapter 3 investigates the contribution of AI to the analysis of flow cytometry data from hematopoietic cells. In *Paper I* we use t-SNE to analyze hematopoietic stem and progenitor cells in comparison to blast and leukemic stem cells from patients with AML and MDS. Further, a method for classifying new samples using t-SNE is developed. In *Paper II*, the uptake of CNDs is included as an additional parameter in the AI-assisted analysis and the selective uptake of CNDs in various healthy and malignant hematopoietic cells is investigated. Chapter 4 extends the research by addressing the questions raised in *Paper II*. In *Paper III,* the selective uptake of CNDs is investigated further and the effect of functionalization is explored. The results regarding the influence of CNDs on lysosomes and autophagy as well as the use of CNDs as carriers for drug transport into the lysosomes are presented in *Paper IV.* Concluding this work, chapter 5 summarizes the findings and discusses potential future research opportunities.

1. Fundamentals

This first chapter sets out the theoretical foundation for understanding the papers that have been published along this thesis. The domains of AI and machine learning are discussed first, followed by an introduction of the physical principles of carbon nanodots. A comprehensive walkthrough of biomedical principles completes this chapter.

1.1 Artificial Intelligence and Machine Learning

The term *Artificial Intelligence* was coined by John McCarthy at a conference at Dartmouth University in 1956 [2], [37]. While the exact definitions of AI vary, it can be understood as a field of computer science that aims to develop intelligent machines and algorithms that mimic human intelligence and which are able to learn and apply the learnings to solve problems [2], [38], [39]. The field of AI encompasses many branches, such as natural language processing, computer vision, and machine learning (ML) [40]. ML enables systems to learn from data without explicit programming [41]. It provides algorithms and methods that allow machines to identify patterns in data to make decisions or predictions based on input data [42]. ML differentiates between supervised, unsupervised and reinforcement learning.

Supervised learning is a process where a model is trained on a labeled dataset [41]. Each input is paired with the corresponding correct output, allowing the model to learn how to map inputs to outputs [43]. This process mimics human learning under the guidance of a teacher [43]. An example of this would be when a child learns what a dog is with the help of its parents. When the child sees a dog, the parents call it a dog so that the child learns over time to recognise a dog independently based on its characteristics and is also able to distinguish it from a cat. To bring it back to ML, the model learns to associate specific features in images with corresponding labels, enabling accurate classification of new, unseen images [41]. Alongside classification another common task in supervised machine learning is regression, enabling models to predict an outcome based on the given input [43].

In contrast, **unsupervised learning** involves unlabeled data, requiring the model to identify patterns or structure without explicit guidance [44]. In the child and parent analogy, the child would learn to distinguish between a dog and a cat in a group of dogs and cats on the basis of the different characteristics of the animals by itself, without knowing the names of the two species. After recognizing the characteristics the child would be able to group the animals into cats and dogs. This would be an example for clustering, which is a common task in unsupervised learning [43]. In clustering, the algorithm groups similar data points together based on specific features, revealing inherent structures in the data [2]. Besides clustering another common task is dimensionality reduction, which are both discussed in detail in the following chapter 1.1.2 and 1.1.1.

Besides supervised and unsupervised learning there is a third basic learning concept namely

reinforcement learning [41]. This type is characterized by learning through feedback from the environment in a trial and error process [43]. An example of this would be training a dog. If the dog behaves as desired, it gets a treat, if not, it is admonished. By rewarding the correct behavior, the dog learns the desired actions, as it wants to maximize positive rewards and avoid negative ones.

1.1.1   Dimensionality Reduction

Dimensionality reduction (DR) is a collection of techniques in unsupervised machine learning and data analysis, which aim to reduce the number of variables or features in a dataset while preserving its essential characteristics [43], [44]. The necessity for the use of dimensionality reduction is frequently evident. In the context of the current era of Big Data, characterized by the growing amount of data, the sheer volume of data produced is increasing continuously, while in the medical field the number of parameters that can be acquired simultaneously is rising in parallel due to technological progress [1], [6], [45]–[47]. Consequently, DR is pertinent to reduce the computing time and the required storage space [45]. Moreover, DR is helpful and sometimes even necessary for visualizing data for analysis, e.g. for revealing hidden features in the dataset, as it becomes more and more challenging to display data given the increasing number of dimensions [4], [7]. If it is possible to reduce a data set with a large number of dimensions to just a few without losing information, the aforementioned challenges can be overcome [44].

DR methods can be subdivided into two main approaches: feature selection (FS) and feature extraction (FE) [41], [44]. While FS selects the smallest possible number of relevant features that contain the information and eliminates the others, FE transforms features of the high-dimensional data set with algebraic transformations into a smaller number of new features for representation in the low-dimensional data set [44], [48]. Since the algorithms used in this work are based on the FE approach, the following paragraphs focus on them. Feature extracting algorithms (FEAs) can further be subdivided into linear or non-linear FEAs. **Linear** dimensional reduction methods, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), seek to identify linear combinations of the original features that capture the greatest variance or discriminative information in the data [44]. These methods assume that the underlying relationships between variables are linear [49]. In contrast, **non-linear** dimensional reduction methods, such as t-Distributed Stochastic Neighbor Embedding (t-SNE), Isometric Mapping (ISOMAP), and Uniform Manifold Approximation and Projection (UMAP) aim to preserve high-dimensional structures of the data by identifying non-linear mappings. Unlike linear methods, non-linear methods are capable of capturing complex relationships and manifold structures present in high-dimensional data [50].
Another difference between the various algorithms can be seen in the preservation of local and global structure. Depending on the design of the algorithm, when creating the low-dimensional mapping, the local or global relationships from the high-dimensional data set are preserved differently [51]. In analogy to our solar system, the local structure would be the distance of the moons to the planets they orbit [51]. The global structure, on the other hand, would reoresent the distances between the individual planets [51].

The following two sections explain the PCA and t-SNE algorithms used in more detail. As PCA is one of the most frequently used and best-known methods, it is only briefly presented while the focus is put on the introduction of t-SNE [49].

4

Principal Component Analysis

Principal Component Analysis (PCA) is an unsupervised, linear dimensional reduction algorithm [52]. In its earliest form, the algorithm goes back to the statistical formulation by Karl Pearson in 1901 [53]. It was then further developed by Harold Hotelling in the 1930s [54]. PCA reduces the feature space to a smaller number of newly created so-called Principal Components (PCs) by determining the directions of the maximum variance in the data set [49]. The new low-dimensional subspace has the PCs as orthogonal axes pointing in the direction of maximum variance [55]. The individual steps of the algorithm are shown in Algorithm 1, which is taken from the review of Anowar et al. [44]. PCA offers several advantages. Primarily, it requires low computation effort since it is not iterative. Additionally, it can reduce over-fitting for training of a model, and it can also be used for noise reduction and data compression [44].

Algorithm 1: PCA

Input: $X \in R^{n \times d}$
Output: $Y \in R^{n \times k}$
1: Construct the covariance matrix $(X.X^T)$
2: Apply linear Eigen decomposition to $(X.X^T)$ to obtain eigenvalues and -vectors
3: Sort Eigen values in decreasing order to sort eigenvectors
4: Build matrix $W(d \times k)$ with k top eigenvectors
5: Transform X using W to obtain the new subspace $Y = X.W$

---

t-Distributed Stochastic Neighbor Embedding

t-Distributed Stochastic Neighbor Embedding (t-SNE) is an unsupervised, non-linear dimensional reduction algorithm presented by Laurens van der Maaten and Geoffrey Hinton in 2008 [21]. The algorithm is a further development of Stochastic Neighbor Embedding (SNE) by Geoffrey Hinton and Sam Roweis [56]. t-SNE enables the visualization of high-dimensional data in a two- or three-dimensional scatterplot, preserving high-dimensional local structure [21]. In order to achieve this, the conditional probabilities $p_{i|j}$ are initially calculated for the high-dimensional distribution between the individual data points, which serve to quantify the degree of similarity between points x_i and x_j.

$$p_{ij} = \frac{exp(-||x_i - x_j||^2/2\sigma_i^2)}{\sum_{k \neq l} exp(-||x_k - x_l||^2/2\sigma_i^2)} \tag{1.1}$$

The step is then repeated for an initially randomly generated, low-dimensional distribution and the conditional probabilities $q_{i|j}$ between y_i and y_j are calculated.

$$q_{ij} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_{k \neq l}(1 + ||y_k - y_l||^2)^{-1}} \tag{1.2}$$

Furthermore, the approach specifies that the conditional probabilities are symmetrical, therefore $p_{ij} = p_{ji}$, $q_{ij} = q_{ji}$ and $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$. Since only pairwise similarities are relevant, p_{ii} and q_{ii} are set to zero.

The difference between the two probability distributions is that a Gaussian distribution is used for the high-dimensional distribution and the Student t-distribution with one degree of freedom is used for the low-dimensional distribution. If the two probability distributions are identical, the low-dimensional distribution will precisely reflect the high-dimensional

distribution. To achieve this, t-SNE employs an iterative process whereby the difference between the two distributions is minimized. The Kullback-Leibler (KL) divergence is employed in the cost function C as a metric for the dissimilarities.

$$C = KL(P||Q) = \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}} \tag{1.3}$$

A gradient method is used for optimisation, the gradient is calculated as follows:

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + ||y_i - y_j||^2)^{-1} \tag{1.4}$$

The gradient can be understood as an attractive or repulsive force between the individual points. The calculation for each point determines how strongly and in which direction the point should be moved. As a result, the t-SNE plot is generated as a function of the calculated coordinates t-SNE1 and t-SNE2 (in the two-dimensional case). Figure 1.1 shows an example of a t-SNE plot.
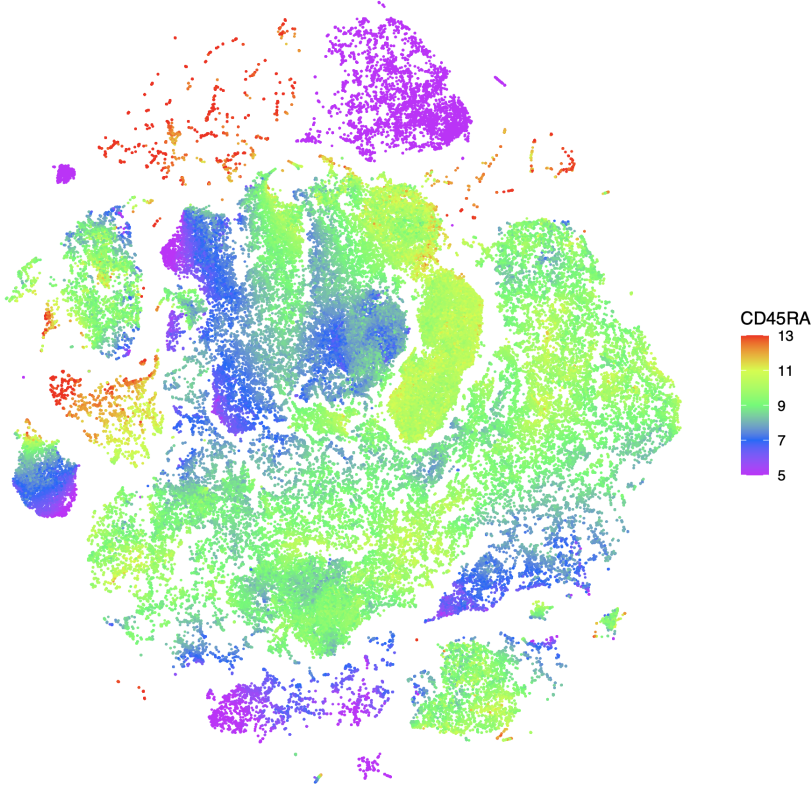

Figure 1.1: *A t-SNE plot of a merged flow cytometry data set from 21 patients is shown as an example. Every point represents a single cell. The fluorescence intensity of the CD45RA marker is colour-coded.*

The main steps of the t-SNE algorithm are summarized in Algorithm 2 based on reference [21].

Algorithm 2: t-SNE

Input: $X \in R^d$

Output: $Y \in R^k$ with k usually 2 or 3

1: Compute pairwise similarities p_{ij} with given perplexity

2: Set $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$

3: Create initial random distribution Y^0 in low-dimensional space

4: Compute pairwise similarities q_{ij}

5: Compute gradient $\frac{\delta C}{\delta y_i}$

6: Adjust Y^t accordingly

7: Repeat step 4 to 6 for the number of given iterations t

The choice of hyperparameters for t-SNE drives the quality of the visualization of the data set [57], [58]. The optimal hyperparameters differ depending on the data set and a large number of studies deal with the choice of hyperparameters [21], [23], [57]–[60]. Of the several hyperparameters available, two are relevant for this work: the number of iterations and the perplexity, which are explained in more detail below. To visualize high-dimensional data, t-SNE optimizes the cost function step by step. The number of steps is determined by the hyperparameter **Number of Iterations** and is to be selected by the user. If the number is set too low, the optimization may end before a stable distribution is achieved [59]. However, a high number of iterations also costs computing time, so a compromise must be found here. Figure 1.2 shows runs of t-SNE at different selected numbers of iterations, with the perplexity parameter fixed. With the number of iterations between ten and 120, the algorithm was stopped before a stable distribution could be achieved.
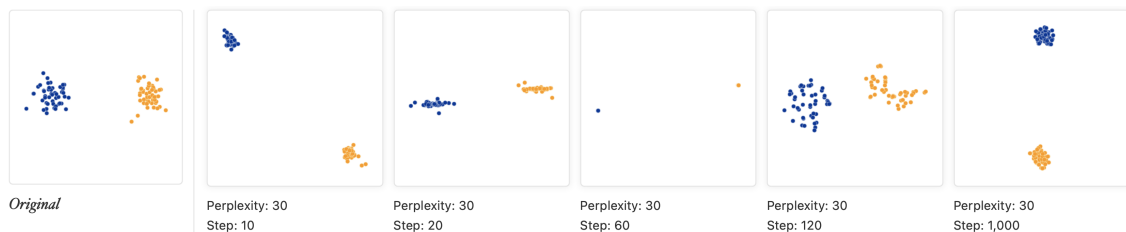

Original   Perplexity: 30   Perplexity: 30   Perplexity: 30   Perplexity: 30   Perplexity: 30
    Step: 10   Step: 20   Step: 60   Step: 120   Step: 1,000

Figure 1.2: *Shown are different t-SNE runs at a fixed perplexity of 30 while the number of iterations is varied between 10 and 1000. The original distribution of the data points is shown on the left. The image was reprinted from [57] and is distributed under a CC-BY 4.0 license by Distill.*

The second hyperparameter, which is also regarded as the most important, is **Perplexity** [23]. The value of the perplexity plays a decisive role in the calculation of the similarities between the points [58]. The selected perplexity influences the width of the Gaussian distribution and thus the number of nearest neighbors of a point [59]. Figure 1.3 shows various runs of t-SNE in which the number of iterations is fixed and the perplexity is varied. The original distribution of the data points can be seen on the left. For small values such as two and five, many small clusters are formed. At five, in contrast to two, there is already a separation between the original clusters. At values of 30 and 50, the two original clusters are easily identifiable. At 100, however, the perplexity seems too high to form separate clusters, as all the points are dispersed again. In this case, the number of data points appears to be smaller than the perplexity.
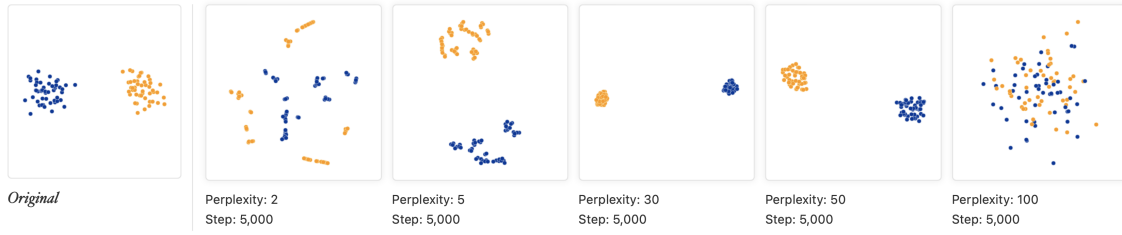
| Original | Perplexity: 2 Step: 5,000 | Perplexity: 5 Step: 5,000 | Perplexity: 30 Step: 5,000 | Perplexity: 50 Step: 5,000 | Perplexity: 100 Step: 5,000 |

Figure 1.3: *Shown are different t-SNE runs at a fixed number of iterations (5000) while the perplexity is varied between two and one hundred. The original distribution of the data points can be seen in the figure on the left. The image was reprinted from [57] and is distributed under a CC-BY 4.0 license by Distill.*

t-SNE is one of the most commonly used algorithms for dimensionality reduction and visualization, but it also has some weaknesses [61]. The nature of t-SNE is to preserve local relationships rather than global structure [21]. This is one of the known limitations of t-SNE [23]. Thus, global structures such as the arrangement of clusters and their distances in the t-SNE plot do not provide a basis for interpretation. Therefore, the axes of the t-SNE plot cannot be interpreted and are omitted, in agreement with common practice. The prior use of a PCA can improve the global structure of the embedding, as suggested by [23], [62]. Another difficulty with the t-SNE algorithm is that, unlike PCA, it cannot add new data points to an existing embedding [59]. This makes it necessary to analyze the entire data set in one run. Most recently, Kobak, D. and Berens, P. [23] as well as Poličar et al. [63] have presented methods to subsequently integrate new samples into an existing embedding.

1.1.2 Clustering

Clustering is a typical task of unsupervised machine learning, in which objects are arranged in groups, so-called clusters, based on the similarities of their features or characteristics [64], [65]. Techniques for cluster analysis can be subdivided into two groups: hierarchical and partitional techniques [66]. In hierarchical cluster algorithms, the objects are placed in hierarchical relationships to each other [65]. The clusters can be formed using a top-down or bottom-up approach, resulting in a dendrogram that maps the relationships between them [64]. Partitional cluster algorithms assign objects to clusters without mapping hierarchical structures [67]. In an iterative process, the data set with n objects is divided into a predefined number i of subgroups by optimizing a criterion function [65]. However, it is often the case that no information about the expected number of clusters is available [68]. So-called automatic clustering algorithms address this problem and refer to all techniques that automatically determine the number of clusters without prior information [65].

1.2 Application of AI in the medical field

The opportunities from the application of AI herald a new era in medicine, with new diagnostic possibilities, improved workflow and research [5]. A distinction is made between two main branches of AI applications: the virtual branch and the physical branch [69]. The virtual branch comprises the area of ML and the applications derived from it, whereas the physical branch includes physical objects, medical devices and robots, e.g. for surgery or care [37]. As the investigations in this thesis relate to the virtual area of applications, examples from this area are given below to illustrate the variety of possible applications.

Diagnostics

AI applications, particularly in the field of ML, have continued to develop in recent years and can help to improve diagnostics for many diseases [41]. An early diagnosis is particularly important in cancer diagnostics. In certain areas, AI can already keep up with experts in terms of accuracy. In the field of dermatology, AI can characterize skin lesions including melanomas as accurately as experts (area under the curve[1] (AUC) 0.94-0.96) (Figure 1.4) [72], [73]. In the interpretation of mammograms in breast cancer screening, the level of experts can also be achieved with an AUC of 0.840 for AI compared to 0.814 for experts [74], [75]. Another example is the use of deep learning (DL, field within ML) to determine molecular status in pathology data. DL can detect and score the expression of a tumor marker protein and predict which tumor tissues have mutations using a neural network model (AUC from 0.733 to 0.856).
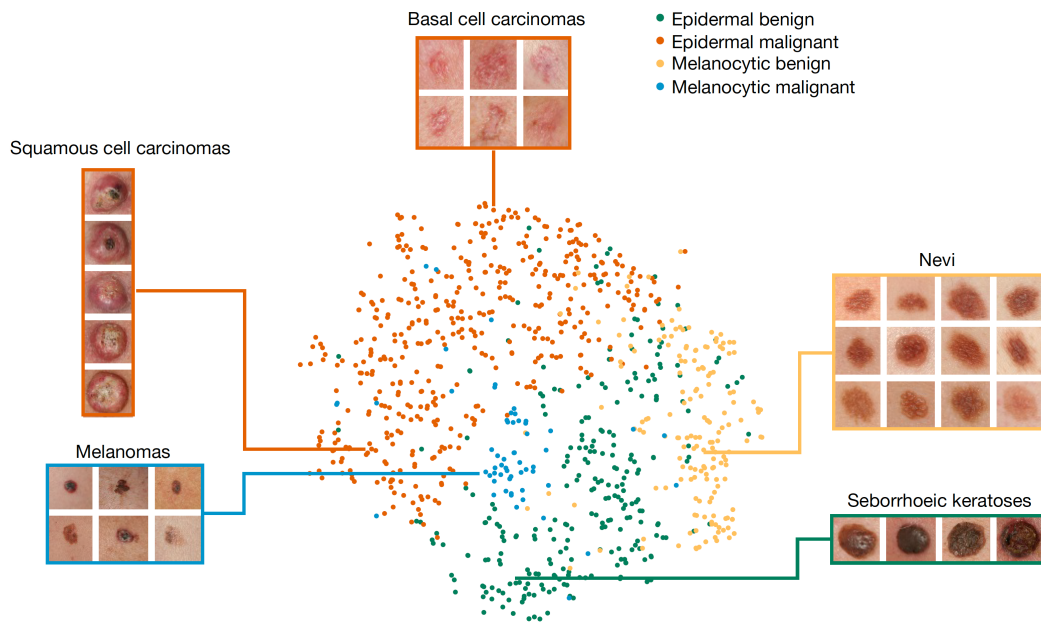

Figure 1.4: *Visualization of the last hidden layer representation of a convolution neural network (CNN) for four disease classes using t-SNE. The four different classes are shown in different colors, revealing the clustering by the algorithm. The image was reprinted from [72]. Reproduced with permission from Springer Nature.*

Precision Treatment

Precision medicine is a new approach of customizing treatment to the patient using multi-module or multi-omics data [4]. This is possible because of the availability of exponentially growing amount of biological and medical data that can be both generated and analyzed by the new technologies [42]. For example, ML models have been developed to predict patient response to treatments using clinical response data [76].
One of the challenges with the existing companion diagnostic assays is that tissue taken from different locations of a tumor may contain different prognostic information due to intra-tumor and inter-tumor heterogeneity [77], [78]. Therefore, another useful application would be the use of AI-based interrogation tools to create a single consistent signature

---

[1]The area under the curve is a measure of the accuracy of a quantitative diagnostic test [70]. A test that is not more accurate than chance has an AUC of 0.5, a perfectly accurate test has an AUC of 1 [71].

from individual tissue slices from a tumor that reflects the entire lesion [79].

Medical Imaging

Another area that particularly benefits from AI is the field of radiology, where a large number of images are used for detection, diagnostics and monitoring of diseases [2]. Many AI applications in this area have already achieved an expert level of accuracy in diagnostics: Lung nodules can be detected on the basis of CT images [80], and pulmonary tuberculosis as well as common lung diseases can be diagnosed with chest radiography [81]–[83]. However, ML methods can not only be helpful to the analysis of medical images, but also with the imaging techniques themselves. For example, DL methods can be used in inter- and intra-modality image synthesis [84] or image segmentation [85].

Drug development

Finding new drugs is challenging and is one of the most difficult parts of drug development due to the enormous number of drug-like molecules that could potentially be synthesized [86]. The application of AI can help in many areas of drug development, such as identifying and validating drug targets, designing new drugs, and improving the effectiveness of research and development [87]. For example, DL was used to identify effective drug candidates and predict their properties and potential toxicity risks [88]. In the field of pharmacogenomics ML and DL methods are being applied which could lead to the drug response being predicted mechanistically, which could be a step towards personalized drug design [89].

1.2.1 AI in flow cytometry

Having discussed examples of the use of AI in the medical field, we will now look in more detail at its use in the field of flow cytometry, which is the main focus of this work. As described in the introduction, the number of parameters that can be measured simultaneously is constantly growing, making manual analysis of the multidimensional data set increasingly complex and time-consuming [90]. Since 2007, efforts have been made to analyze the data sets using computational methods in order to overcome the difficulties of manual analysis [91]–[97]. However, as the results were initially difficult to compare because different data sets and evaluation methods were used when testing the algorithms, the *Flow Cytometry: Critical Assessment of Population Identification Methods* (FlowCAP) project was set up to solve these problems [97]. The aim was to test and compare the algorithms for specific tasks using the data sets provided [97]. The results of the first two FlowCAP challenges are presented in the publication by Aghaeepour et al. [97] and are briefly summarized below, followed by a short presentation of the results from FlowCAP III and IV.

The FlowCAP I was about comparing the results from computational tools with the results from manual gating carried out by experts. There were four different categories in which the algorithms could compete: 1) completely automatic, 2) manually tuned, 3) assignment of cells to populations with predefined numbers of populations, and 4) supervised approaches trained with human-provided gates. Within the four categories, the algorithms were tested on five different data sets. The F-measure statistic was used for the comparison, a value of 1.0 means that the result of the manual gating could be reproduced exactly. Many algorithms achieved good results in several challenges with an F-measure

above 0.85. Some algorithms were among the top, e.g. ADICyt in categories 1-3 and SamSPECTRAL in challenge 3, while others were among the top group only for certain data sets, such as flowMeans, FLOCK and FLAME in category 1.

In FlowCAP II, the focus was on the use of biomarker patterns in the data sets to classify the samples. The algorithms were tested using three different data sets for different diseases. The values for precision, recall, accuracy and F-measure were used for evaluation. It was found that in two of the three data sets, many algorithms were able to perform the classification perfectly (F-measure = 1.0). In the third data set, none of the algorithms performed well, which in retrospect was probably due to the fact that no unambiguous classification could be performed using the markers used.

FlowCAP III focused on the identification of rare cells using computational methods [98]. Training samples were provided, each data set contained two populations of rare cells, which made up approximately 0.02 % to 0.04 % of the total population [98]. The challenge was to identify the rare cells using computational methods. The accuracy of the identification was determined by comparison with manual gating [98]. In paper [98], Peng Qiu presents the method that achieved the highest accuracy in the challenge with an F-measure of 0.64 in phase-one and 0.69 in phase-two.

The FlowCAP IV challenge was about establishing a benchmark for assessing the relative performance of algorithms in identifying cellular correlates of clinical outcome [99]. The results were published by Aghaeepour et al. in paper [99]. A key finding was that automated, unbiased approaches that include the expression of all markers are preferable, as they do not miss any potentially relevant cell populations. Furthermore, it was shown that the correlates of biologically important outcomes can be based on cell types that are not well described in the literature.

The FlowCAP challenges showed that automated methods can keep up with or are already superior to manual methods for many applications. Nevertheless, a study from 2021 showed that half of the laboratories were still not using automated methods for cell population identification [90]. However, development is continuing as review articles on current trends in the field of automated analysis show [18], [19], [90], [100]–[102]. Some examples from recent years are summarized below:

Zhong et al. [103] have introduced an AI-supported workflow for the diagnosis of acute leukemia for faster and more objective identification of malignant cell populations. The AI model used is mainly based on DR, clustering and classification algorithms. Good agreement was achieved between the manual results and the results using AI (Pearson correlation coefficient of 0.913 ($p < 0.04$)).

Aanei et al. [104] investigated differences between AML blasts and normal myeloid progenitor cells using high-dimensional analysis algorithms (including Citrus, viSNE, SPADE). The algorithms were used to identify the most informative markers for differentiation within a panel to be analyzed.

Cheng et al. [105] have used deep learning to train AI for cell classification and detection of acute leukemia. The application to the acute leukemia orientation tube (ALOT) protocol established by Euroflow was investigated. A sensitivity of 94.6 % was achieved for

the detection of AML and 98.2 % for B-lymphoblastic leukemia (B-ALL) patients. While the sensitivity for the detection of pathological cells varied, it was at least 80 % for physiological cells.

Clichet et al. [106] have developed an ML model that generates a prediction score based on five parameters for the diagnosis of high- and low-risk MDS. The sensitivity is 91.8 % with a specificity of 92.5 %.

Vial et al. [107] and Salama et al. [108] both applied AI for the automated analysis of multiparameter flow cytometry data to identify minimal residual disease (MRD). Vial et al. were thus able to identify the MRD in AML using the unsupervised algorithm Flow-Self-Organizing-Maps (FlowSOM). The results showed good agreement with the molecular MRD data, which could enable the diagnosis of MRD in AML patients without molecular markers. Salama et al. developed deep neural networks (DNNs) and were able to recognize MRD in chronic lymphocytic leukemia (CLL) using a hybrid approach.

Above mentioned research provides strong indication that the use of AI can significantly contribute to the analysis of flow cytometry data, diagnosis and identification of MRD, showing the high relevance for further investigations.

1.3 Carbon nanoparticles

As this thesis includes the investigation of the uptake and influence of carbon nanodots (CNDs) on cellular metabolic processes in *Paper II*, *Paper III* and *Paper IV*, the physical basis for this will be presented in the following chapter. Additionally, the uptake of the CNDs as parameters for visualization in a t-SNE plot is examined in Paper II. In the first section, low-dimensional carbon modification graphene and its physical properties are introduced. Afterwards, graphene quantum dots are reviewed. Finally, CNDs and their biomedical applications are discussed.

1.3.1 Graphene

Graphene is a allotrope of the element carbon. Novoselov and Geim succeeded first in preparing and investigating this two-dimensional modification in 2004 [109]. Until then, it was assumed that infinitely extended, two-dimensional structures were thermodynamically unstable according to the Mermin-Wagner theorem and therefore unrealisable [110]. The two physicists were awarded the Nobel Prize in Physics in 2010 for their discovery.

In graphene, the sp^2-hybridized carbon atoms are arranged in a hexagonal lattice with a diatomic base. This results in the two sublattices A and B (Figure 1.5). The bond lengths are each a = 1.42 $\mathring{A}$ with a bond angle of 120°. This results in a lattice constant of around $246\,pm$. The reciprocal lattice of graphene also has a hexagonal structure. The vertices of the first Brillouin zone, the so-called Dirac points (K, K'), play an important role in understanding the unusual band structure of graphene. Using the tight-binding model (TBM), the band structure of graphene can be calculated as in [111], which allows a precise observation at the Dirac points:

The real lattice vectors can be written as

$$\vec{a_1} = \frac{a}{2}(3, \sqrt{3}), \ \vec{a_2} = \frac{a}{2}(3, -\sqrt{3}) \tag{1.5}$$

and the reciprocal lattice vectors as

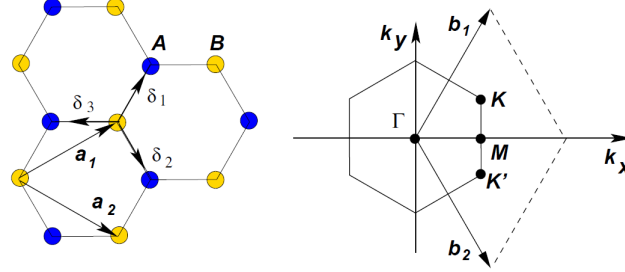$$\vec{b_1} = \frac{2\pi}{3a}(1, \sqrt{3}),\ \vec{b_2} = \frac{2\pi}{3a}(1, -\sqrt{3}) \tag{1.6}$$

Figure 1.5: *The lattice structure of graphene is shown with the lattice vectors a_i and the vectors to the nearest neighbors δ_i. The sublattice A (B) is shown in blue (yellow). On the right is the corresponding 1st Brillouin zone of graphene with the reciprocal lattice vectors b_i and the K, K', M and Γ points. The image was reprinted from [111] ©(2009) by the American Physical Society.*

The tight-binding Hamiltonians for the electrons in the graphene, taking into account that they can jump to the nearest neighbor as well as to the next nearest neighbor, can be represented as:

$$H = -t \sum_{\langle i,j\rangle,\sigma} (a^{\dagger}_{\sigma,i} b_{\sigma,j} + h.c.) - t' \sum_{\langle i,j\rangle,\sigma} (a^{\dagger}_{\sigma,i} a_{\sigma,j} + b^{\dagger}_{\sigma,i} b_{\sigma,j} + h.c.) \tag{1.7}$$

with t ($\approx 2.7\,eV$) as the hopping energy to the nearest neighbor (between the sublattices) and t' ($\approx 0.1\,eV$) as the hopping energy to the next nearest neighbor (within a sublattice). With $a_{\sigma,i}(a^{\dagger}_{\sigma,i})$ the annihilation (creation) operator for an electron with the spin σ in the sublattice A is denoted. The same applies to $b_{\sigma,i}(b^{\dagger}_{\sigma,i})$ with respect to sublattice B. The energy dispersion derived from this can be described by

$$E_{\pm}(\vec{k}) = \pm t\sqrt{3 + f(\vec{k})} - t'f(\vec{k})$$
$$f(\vec{k}) = 2cos\,(\sqrt{3}k_y a) + 4cos\,(\frac{\sqrt{3}}{2}k_y a)\,cos\,(\frac{3}{2}k_x a) \tag{1.8}$$

The plus respective minus sign refers to the binding (π) respective anti-binding (π^*) band. For $t' = 0$ the spectrum is symmetrical around the zero energy. For finite values of t', the electron-hole symmetry breaks down and the π- and π^*-band become asymmetric. Figure 1.6 shows the band structure of graphene for t and t'. By developing the band structure (1.8) in the neighborhood of K (K') with $\vec{k} = \vec{K} + \vec{q}$ and $|\vec{q}| \ll |\vec{K}|$ it results in:

$$E_{\pm}(\vec{k}) \approx \pm v_F |\vec{q}| + \mathcal{O}((\vec{q}/K)^2) \tag{1.9}$$

Due to the linear dispersion (1.9) near the Dirac points, the charge carriers there behave like relativistic Dirac fermions with the Fermi velocity $v_F = 1 \times 10^6 m/s$ [111]. The equation of motion of the charge carriers is therefore subject to the Dirac equation:

$$- iv_F \sigma \nabla \Psi = E\Psi \tag{1.10}$$

13

with the Pauli matrices $\sigma = (\sigma_x, \sigma_y)$, the energy eigenvalue E and the two-component wave function $\Psi = (\psi_A, \psi_B)$ for the sublattices A and B.
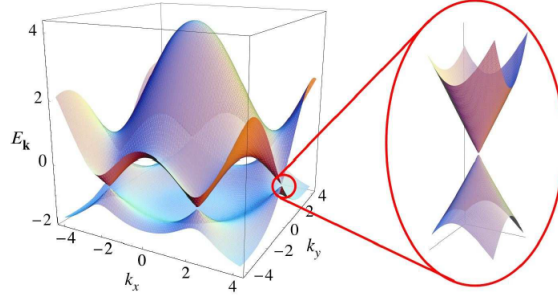

Figure 1.6: *Schematic representation of the band structure of graphene in the vicinity of the Dirac points. The enlarged view shows that the band gap vanishes at the Dirac points. The image was reprinted from [111] ©(2009) by the American Physical Society.*

The band gap between the valence and conduction band vanishes in the Dirac points (Figure 1.6). Graphene is therefore a semiconductor with a vanishing band gap. As a result, interesting transport phenomena such as the quantum Hall effect at room temperature or the Klein paradox can be observed [111].

1.3.2 Graphene quantum dots

Graphene quantum dots (GQDs) are zero-dimensional graphene derivatives consisting of single to a few layers of graphene with a diameter of a few nanometers [112]. The decisive difference between the material graphene and GQDs is the band gap. While graphene is a material with a vanishing band gap, GQDs have a band gap due to quantum confinement effects and edge effects [113]. The size of the band gap seems to be influenced by the number of aromatic carbon rings and thus by the size of the GQDs. As the number of rings increases, the band gap decreases. Eda et al. [114] were able to show this using simulations based on density functional theory. The experimental results of Kim et al. [115] and Ritter et al. [116] are consistent with their findings. On the other hand, influences of the band gap by the edges of the GQDs can be observed. Both the modification of the edges by nitrogen-containing functional groups and the ratio of the possible edge types of graphene *zigzag* and *armchair* have an effect on the band gap of the GQDs [115]–[117]. If the number of attached amino groups increases, the band gap decreases [117].

The changes in the band gap have a direct effect on the fluorescence of the GQDs, which extends from the UV to the visible range [118]. However, the origin of the fluorescence has not yet been fully explained [112]. In contrast to II/VI semiconductor quantum dots, the fluorescence wavelength is not a monotonic function of the particle size [112]. This indicates a subordinate dependence on the size quantization. The pH value of the environment also has an effect on the fluorescence intensity. The intensity increases with increasing pH value [119].

1.3.3 Carbon nanodots

The GQDs discussed above are a pristine system. In the synthesis of carbon nanoparticles for biomedical applications, however, we are often no longer dealing with a pure system due

to the possible presence of heteroatoms, mixed edge structures and various hybridization forms of carbon. These forms of heterogeneous systems as well as GQDs are summarized in this work under the term carbon nanodots (CNDs). In the following, the synthesis and modification methods are first described and then the physicochemical properties of the CNDs are discussed.

Synthesis and modification

Two approaches can be distinguished when synthesizing carbon nanoparticles: the *top-down* approach and the *bottom-up* approach [118]. In the top-down approach, the nanoparticles are obtained by chemical, electrochemical or physical methods from the cutting of carbon materials [113]. The most common method is cutting with concentrated oxidizing acid [118]. Other methods involve hydrothermal/solvothermal/special oxidation [120], [121], electrochemistry [122], [123], metal-graphite intercalation [124], arc discharge [125], laser ablation [126] and nanolithography with reactive ion etching [127], [128]. In contrast, in the bottom-up approach, they are synthesized from organic molecules or polymers [118]. During synthesis, -COOH, -OH, -NH$_2$ and other functional groups form covalent bonds under dehydration [129]. Dehydration and carbonization can be achieved by a variety of approaches such as hydrothermal [130], microwave [131] and combustion [132] methods, pyrolysis in concentrated acid, [133] and carbonization in a microreactor [134].

The surface groups of the CNDs vary according to the precursors used and the reaction conditions. Typically, carboxyl, hydrophilic hydroxyl or amino groups decorate their surface, providing a target for covalent modifications [135]. Other substances such as polymers, DNA, proteins or antibodies can be attached to the CNDs through covalent bonds with the functional groups [136], [137]. The most commonly used modification method is the amide coupling reaction in which 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and N-hydroxysuccinimide (NHS) are used as coupling reagents [136]. In the modifications of the CNDs carried out in *Paper III* and *Paper IV*, this method was used to attach amino sugar, amino terminated sugar oligomers or branched polyethyleneimine (bPEI) to the CNDs. Another type is non-covalent modification by means of electrostatic, complexation or π-interactions [129], [135].

Fluorescence properties

As described above, there are a large number of different synthesis methods and modification options. The optical properties of the CNDs can vary depending on the synthesis method or modification used [129]. This makes it difficult to gain a comprehensive understanding of the relationship between the optical properties and the structure of the CNDs [138]. In the following, five mechanisms that are most frequently considered to explain the fluorescence mechanism of CNDs are briefly presented.

i.) The first effect is the **quantum confinement** effect of the conjugated π-domains. The size of the sp^2-hybridized domain has a decisive influence on the fluorescence [114]. The radiative recombination of electron-hole pairs in such domains can generate fluorescence [118].

ii.) The **surface/edge state** influences the fluorescence emission of the CNDs. The surface state is determined by the hybridization of the carbon skeleton and the associated chemical groups [129]. Carboxyl, carbonyl and amide groups are mainly

responsible for the green emission, while hydroxyl groups are accountable for the blue emission [139], [140].

iii.) In non-conjugated systems such as polymeric dots, the **crosslink-enhanced emission** effect plays an important role [138]. For cross-linked polymer chains, vibrations and rotations are reduced, resulting in increased fluorescence emission intensity [141], [142].

iv.) Another mechanism is based on the **self-trapping of an exciton** in a polycyclic aromatic hydrocarbon (PAH) network [143]. This results in a considerable stokes shift and selective excitation of the various subsets of PAHs [143].

v.) The last mechanism is based on the excitation of **small molecular fluorophores**, which are covalently or non-covalently bound to the CNDs [144]. This is particularly relevant for bottom-up synthesized CNDs that use citric acid as a precursor [145], [146].

Due to the structural heterogeneity of the individual CNDs, the origin of the fluorescence may be a combination of the different mechanisms [138]. The quatum yield (QY) is the ratio of the number of photons emitted to the number of photons absorbed. The QY achieved vary greatly for different CNDs and can range from less than 1 % to over 90 % [118], [146].

The fluorescence of CNDs can be used to detect their uptake into cells, for example using flow cytometry (section 2.1). The fact that CNDs are generally taken up by cells has already been shown in studies [147]–[149]. As the toxicity of CNDs appears to be low compared to semiconductor quantum dots, they have a high water solubility, are biocompatible, can be functionalized and have a fluorescence that can be influenced, they open up a wide range of possibilities for biomedical applications, which are explored in the next paragraph [113], [148], [150].

1.3.4 Biomedical application of carbon nanoparticles

Due to the ease of production, the influenceable optical properties and the possibility of functionalisation with various ligands and biomolecules, carbon nanoparticles offer a whole spectrum of possible applications in the field of biomedicine [25], [151]. In the following, possible fields of application will be presented.

Intracellular sensing

One field of application is intracellular sensing, in which the nanoparticles are used as nanosensors in vitro or in vivo [25]. Interactions with the environment can lead to changes in the fluorescence properties of the nanoparticles, the measurement of which can be used to infer changes in the physical property of interest, such as the pH value [25]. There are already a large number of applications for measuring the pH value [152]. Wu et al. [26] have presented N-doped GQDs whose fluorescence intensity shows a linear dependence on the pH value for values between 2 and 9. The authors explain this by the dependence of the zeta potential on pH. The protonation and deprotonation of the functional groups on the GQD surface leads to changes in the zeta potential, which exhibits a linear dependence on the pH value. Nie et al. [27] synthesized and functionalized carbon dots in which the ratio of fluorescence intensity at different peaks changes linearly with pH for values between 5

and 8. Other areas of application include the detection of ions or reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2) [25]. Heavy metals can accumulate in organs or tissues and cause a wide range of diseases [153]. The concentrations of mercury ions (Hg^{2+}) [28], tin ions (Sn^{2+}) [154] or iron ions (Fe^{3+}) [155] can be detected by the quenching of the fluorescence caused by complex formation between the ions and the CNDs. ROS can also be detected by fluorescence quenching. Shan et al. [29] developed boron doped carbon quantum dots (CQDs) in which quenching occurs in the presence of ROS through charge transfer between the doped boron atoms and hydrogen peroxide.

Drug delivery

One disadvantage of conventional drug delivery systems is that they can have a number of undesirable side effects such as incorrect dosing due to low specificity, which can lead to increased cellular toxicity with serious consequences for the patient [25]. Carbon nanoparticles appear to be suitable candidates to enable risk-reduced drug delivery as nanocarriers due to their good water solubility, high biocompatibility, low toxicity and tunable optical properties [24], [25]. Wang et al. [30] explored hollow carbon dots as nanocarriers for the anticancer drug doxorubicin (DOX). DOX binds to the carbon dots by means of interactions such as $\pi - \pi$ stacking and electrostatic interactions [30]. A rapid uptake by the cells as well as a pH-controlled release of DOX in vitro could be demonstrated. Furthermore, no influence on drug activity was observed. Tang et al. [31] were also able to establish a drug delivery system in vitro for DOX using CNDs, in which the drug release can be observed in real time using Förster resonance energy transfer (FRET). The release of DOX is pH-dependent. In both publications, the maximum release was in the rather acidic pH range of 5-5.5. This could play an important role in cancer treatment, as the microenviroment of tumors is often more acidic compared to healthy tissue [156], [157].

Antimicrobial Therapy

The global increase of bacterial resistance to antibiotics is causing growing concern worldwide [158]. Certain bacterial infections are resistant to almost any antibacterial treatments currently available [159]. This calls for new antimicrobial treatment methods. Antimicrobial photodynamic inactivation (PDI) appears to be a promising method for inactivating many microorganisms [158], [160]. The basis of PDI is the non-specific oxidative damage to biomolecules in the cell membrane or within the cells by ROS [160]. Photoexcited CNDs can generate ROS, as has been shown in various studies, making them promising candidates as antimicrobial agents [33]–[36]. Based on the previous results, electron/hole pairs (e^-/h^+) are formed in the photoexited CNDs [159]. Both e^- and h^+ can lead to the formation of ROS through charge transfer to surrounding substrates, which is referred to as *Type I* photosensitization [158], [160]. In *Type II*, there is a direct energy transfer with ground-state molecular oxygen ($^3O_2$), which leads to the formation of singlet oxygen ($^1O_2$) [158], [160].

1.4 Biomedical fundamentals

This chapter discusses biological and medical principles relevant to this thesis. The heart of the work, namly *Paper I* and *Paper II*, deals with the hematopoietic system, the various blood cells and malignant diseases of the hematopoietic system. Therefore, the next section delineates the process of healthy blood formation alongside diverse blood cell types. This section is followed by an examination of the specific markers utilized for distinguishing

these blood cells. The subsequent section addresses the implications of malfunction in the blood-forming system, focusing particularly on the diseases AML and MDS. Important cellular processes of endocytosis and autophagy are explored in the subsequent, too. These processes are relevant to understand the mechanisms of the nanoparticle uptake into and the distribution inside the cell discussed in *Paper II*, *Paper III* and *Paper IV*.

1.4.1 Hematopoiesis

Hematopoiesis is the process of blood cell formation. During fetal development up to the 5th month, hematopoiesis occurs in the liver and spleen [161]. After that, blood cells mature exclusively in the bone marrow (BM). The multipotent hematopoietic stem cells (HSCs), which are capable of self-renewal, reside in the BM and are the origin of hematopoiesis [162]. The differentiation process is regulated by numerous hematopoietic growth factors. Thanks to research, the understanding of hematopoiesis is constantly increasing, so that the model has continued to develop over the years, as can be seen in Figure 1.7 [163].
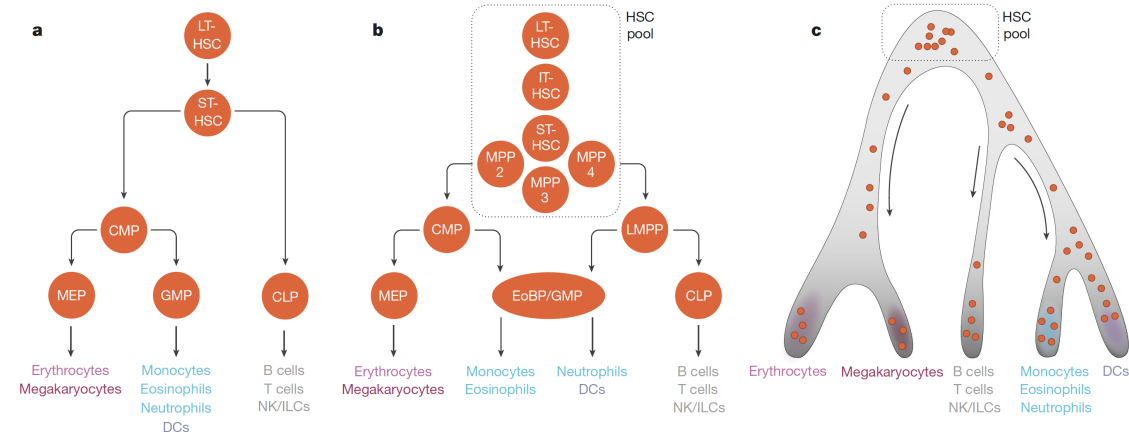

Figure 1.7: *Visualisation of the different hierarchical models of hematopoiesis. a) Visualisation based on the research around 2000. Hematopoietic stem cells (HSCs) include the long-term (LT) HSCs and the short-term (ST) HSCs, which only have a limited potential for self-renewal. The common myeloid progenitor cells (CMP) and common lymphoid progenitor cells (CLP) develop from the ST-HSCs. The CMPs branch into megakaryocyte-erythrocyte progenitor cells (MEPs) and granulocyte-monocyte progenitor cells (GMPs). b) Between 2005 and 2015, the HSC pool became more heterogeneous and also contains multipotent progenitor cells (MPPs). The myeloid and lymphoid branches remain connected even longer via the lymphoid-primed multipotential progenitor cells (LMPPs). c) From 2016 onwards studies suggest that development may be more of a continuous process rather than a discrete hierarchy. The image was reprinted from [163]. Reproduced with permission from Springer Nature.*

In the classical model of hematopoiesis, the functional long-term HSCs (LT-HSCs) evolve into short-term HSCs (ST-HSCs), which have only a limited potential for self-renewal and can differentiate into the multipotent progenitor cells (MPPs) [164]. The MPPs differentiate into the various precursor cells, on the one hand the myeloid lineage, commen myeloid progenitor cells (CMPs), and on the other hand the lymphatic lineage, common lymphoid progenitor cells (CLPs) [165]. The further developmental stages from the CMPs are the granulocyte-macrophage (GMPs) and megakaryocyte-erythrocyte progenitors (MEPs) [163]. CLPs give rise to T-, B-, NK- and dendritic cells, while GMPs differentiate into granulocytes and monocytes; MEPs evolves into megakaryocytes and erythrocytes [165].

However, the classical model simplifies the complexity of the HSCs and the precursor cells and has been further developed over the years [165]. Various studies in recent years suggest that such development may be more of a continuous process rather than a discrete hierarchy [166]–[168].

The following section describes the composition and function of the blood and the different blood cells in more detail:

Blood undertakes a multitude of diverse transport and communication tasks [169]. Oxygen, nutrients, and hormonal messengers are transported by the blood to the organs, which in turn release metabolic by-products and hormones into the blood [161]. To maintain the system, it is necessary to maintain the blood volume as well as its components nearly constant [161]. Blood comprises approximately 55% plasma, which in turn consists of 91.5% water with dissolved molecules [170]. The remaining 45% constitutes blood cells, which can be divided as follows:

Erythrocytes, also known as red blood cells, constitute about 95% of cells in the blood volume [169]. They are primarily responsible for transporting oxygen from the lungs to other organs and for transporting carbon dioxide back [171]. The responsible transport protein is hemoglobin, which imparts the blood its characteristic red color [171]. Erythrocytes are biconcave disc-shaped with a mean diameter of $7-8\,\mu m$ and lack a nucleus [170]. They have a lifespan of approximately 120 days [171].

Platelets are also nucleus-free and biconcave in shape with a diameter of $1-4\,\mu m$. The normal count ranges from 170.000 to 400.000 per μl of blood, where they initially exist in an inactive state [161]. Platelets are crucial for blood clotting and wound healing, a process known as hemostasis [161]. In the event of vascular injury, they bind to components of the subendothelium, initiating their activation [161]. Consequently, they release platelets and coagulation-activating mediators, leading to the formation of a platelet aggregate [171].

Leukocytes, also known as white blood cells, are a vital component of the immune system [161]. As they can actively move, they are capable of leaving the blood vessels and migrating into tissues [171]. Essentially, they are categorized into granulocytes, lymphocytes, and monocytes: Granulocytes belong to the non-specific innate immune system [161]. They possess a lobed or C-shaped nucleus and numerous granules[2] [171]. Lymphocytes further divide into B-lymphocytes (10%), T-lymphocytes (80%), and natural killer (NK) cells (10%) [171]. B- and T-lymphocytes are part of the specific adaptive immune system, while NK-cells belong to the non-specific part [161]. Monocytes transform into macrophages in tissues [170]. These macrophages destroy bacteria and foreign bodies through phagocytosis [171].

1.4.2 CD antigens

Blood cells can not completely be morphologically distinguished from one another [172]. However, depending on their differentiation or state of activity, they express a characteristic profile of surface molecules [173]. This is schematically depicted in Figure 1.8 for T- and B-cells. These surface molecules allow for the distinction of various (sub-)populations of blood cells [174]. Hence, these surface molecules are also referred to as differentiation

---

[2]Granular structures present in the cytoplasm **BiofuerMed**.

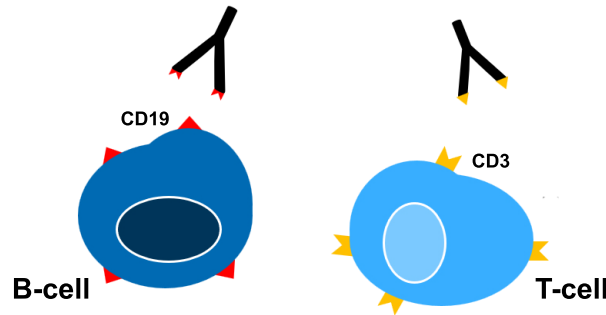markers or cluster of differentiation (CD) antigens [173].

Figure 1.8: *Schematic representation of CD antigens with the corresponding CD antibodies using the example of T- and B-cells. T-cells express CD3 antigens and B-cells express CD19 antigens to which the corresponding antibodies can bind.*

To identify the different surface molecules, specific monoclonal antibodies are utilized [173]. These antibodies can only bind to a specific molecular differentiation antigen [172]. The resulting antibody bindings can then be visualized using various methods. For example, it is possible to label the antibodies with fluorochromes [173]. This is exploited, among other techniques, in flow cytometry, which is discussed in more detail in Section 2.1. The so-called CD system is utilized for immunophenotyping [173].

1.4.3 Hematological diseases

This section discusses the diseases of the hematopoietic system that are relevant to this work, namely acute myeloid leukemia (AML) and Myelodysplastic neoplasms (MDS).

Leukemia

Leukemia is a malignant disease of hematopoietic cells characterized by uncontrolled proliferation of immature hematopoietic precursor cells, so called blasts [171]. This leads to displacement of normal hematopoiesis, with pathological cells being released into the blood [175]. Acute and chronic forms are distinguished, with the acute form having a rapidly progressive and highly aggressive course, often leading to death within a few weeks [175]. A diagnosis of acute leukemia is made if more than 20%[3] lymphoblasts or myeloblasts are present in the bone marrow at the time of diagnosis [169], [176]. If there is a proliferation of lymphatic precursor cells of the BM, the lymphatic system or the thymus, the disease is referred to as acute lymphoblastic leukemia (ALL), if there is a proliferation of myeloid precursor cells, the disease is referred to as AML [175]. If the disease cannot be clearly assigned to either of the two lines, it is referred to as acute leukemia without a clear lineage [175].
AML accounts for 75-80% of acute leukemias in adults, whereas in children, it represents only 15-20% [175]. Known risk factors include genetic diseases, ionizing radiation, specific toxins, cytostatic drugs, and pre-existing conditions [171]. Instead of mature and functional blood cells, AML leads to the development of non-functional blasts. Consequently, symptoms such as general malaise, weakness, pallor, fatigue, bruising, petechiae (pinpoint-sized skin and mucous membrane bleedings), nosebleeds, gum bleeding, and fever may occur [175]. Treatment options vary depending on the type of leukemia and the individual

---

[3]AML with defined genetic abnormalities may be diagnosed with less than 20% blasts [176].

risk profile of the patient, ranging from chemotherapy and stem cell transplantation to further specific therapies [175]. The treatment goal is molecular/cytogenetic remission. Without treatment, the majority of AML patients succumb within three months. Thanks to modern treatment modalities, 15-70% (depending on the risk group) of younger patients survive five years after, whereas older patients > 60 years have a significantly worse prognosis, with a 5-year survival rate of 10-15% [175]. Common challenges include frequent relapses and the emergence or pre-existence of resistance to various drugs (multidrug resistance) [171], [177]–[179].

Table 1.1: *Classification of Acute myeloid leukemia (AML). Reproduced from [176], which is distributed under a CC BY 4.0 license by Springer Nature.*

Acute myeloid leukaemia with defining genetic abnormalities
Acute promyelocytic leukaemia with PML::RARA fusion
Acute myeloid leukaemia with RUNX1::RUNX1T1 fusion
Acute myeloid leukaemia with CBFB::MYH11 fusion
Acute myeloid leukaemia with DEK::NUP214 fusion
Acute myeloid leukaemia with RBM15::MRTFA fusion
Acute myeloid leukaemia with BCR::ABL1 fusion
Acute myeloid leukaemia with KMT2A rearrangement
Acute myeloid leukaemia with MECOM rearrangement
Acute myeloid leukaemia with NUP98 rearrangement
Acute myeloid leukaemia with NPM1 mutation
Acute myeloid leukaemia with CEBPA mutation
Acute myeloid leukaemia, myelodysplasia-related
Acute myeloid leukaemia with other defined genetic alterations
Acute myeloid leukaemia, defined by differentiation
Acute myeloid leukaemia with minimal differentiation
Acute myeloid leukaemia without maturation
Acute myeloid leukaemia with maturation
Acute basophilic leukaemia
Acute myelomonocytic leukaemia
Acute monocytic leukaemia
Acute erythroid leukaemia
Acute megakaryoblastic leukaemia

AML can be classified according to the WHO classification [176] or the International Consensus Classification (ICC) [180], which were updated in 2022. An older classification is the FAB classification (French-American-British working group of leukemia) [181]. The FAB classification categorizes AML into subtypes M0 to M7 based on cytomorphological and cytochemical criteria. Meanwhile, the WHO classification and ICC classification incorporate immunological, cyto-, and molecular genetic changes, providing a more comprehensive classification scheme [176], [180]. According to the 2022 WHO classification, AML can be divided into AML with defining genetic abnormalities and AML defined by differentiation [176]. The different subtypes are shown in table 1.1.

Myelodysplastic neoplasms

Myelodysplastic neoplasms (MDS) comprise a heterogeneous group of stem cell diseases characterised by impaired proliferation and maturation of hematopoietic cells [182]. In the 2022 WHO classification [176], the myelodysplastic syndromes were renamed myelodysplastic neoplasms, but the previously known abbreviation "MDS" remains.

Table 1.2: *Classification and defining features of myelodysplastic neoplasms. Reproduced from [176], which is distributed under a CC BY 4.0 license by Springer Nature.*

MDS with defining genetic abnormalities	Blasts	Cytogenetics	Mutations
MDS with low blasts and isolated 5q deletion (MDS-5q)	< 5% BM and < 2% PB	5q deletion alone, or with one other abnormality other than monosomy 7 or 7q deletion	
MDS with low blasts and SF3B1 mutation[a] (MDS-SF3B1)	< 5% BM and < 2% PB	Absence of 5q deletion, monosomy 7, or complex karyotype	SF3B1
MDS with biallelic TP53 inactivation (MDS-biTP53)	<20% BM and PB	Usually complex	Two or more TP53 mutations, or 1 mutation with evidence of TP53 copy number loss or cnLOH
MDS, morphologically defined			
MDS with low blasts (MDS-LB)	< 5% BM and < 2% PB		
MDS, hypoplastic[b] (MDS-h)	< 5% BM and < 2% PB		
MDS with increased blasts (MDS-IB)			
MDS-IB1	5–9% BM or 2–4% PB		
MDS-IB2	10-19% BM or 5–19% PB or Auer rods		
MDS with fibrosis (MDS-f)	5–19% BM; 2–19% PB		

[a] Detection of $\geq$ 15% ring sideroblasts may substitute for SF3B1 mutation. Acceptable related terminology: MDS with low blasts and ring sideroblasts.
[b] By definition,$\leq$ 25% bone marrow cellularity, age adjusted.
BM bone marrow, PB peripheral blood, cnLOH copy neutral loss of heterozygosity.

MDS can be classified according to the WHO classification [176] or the International Consensus Classification (ICC) [180], which were updated in 2022. According to the new WHO classification, MDS are divided into MDS with defined genetic abnormalities and MDS that are morphologically defined [176]. An overview is depicted in Table 1.2.

The severity of the disease can vary from indolent conditions with an almost normal life expectancy to forms that are similar to acute myeloid leukemia [183]. MDS can develop into AML if the blast threshold of 20% is exceeded [182]. The difference between MDS and non-neoplastic clonal myeloid proliferations is the presence of morphological dysplasia [184]. MDS is one of the most common hematological systemic diseases affecting adults [185]. The incidence is age-dependent with 3-5 out of 100.000 per year in the general population and up to 40 out of 100.000 per year in over 70-year-olds [185]. As with AML, known risk factors include genetic diseases, ionizing radiation, specific toxins, cytostatic drugs, and pre-existing conditions [185]. Due to ineffective hematopoiesis and the resulting cytopenia, symptoms such as fatigue, reduced performance, tachycardia, dyspnoea on exertion, epistaxis, petechiae, haemorrhages, haematomas, fever, sepsis, pneumonia, recurrent infections and fungal infections may occur [182], [185]. Whether and which therapy is necessary depends on various individual factors [185]. Treatment options include several drug therapies, chemotherapy and stem cell transplants [185].
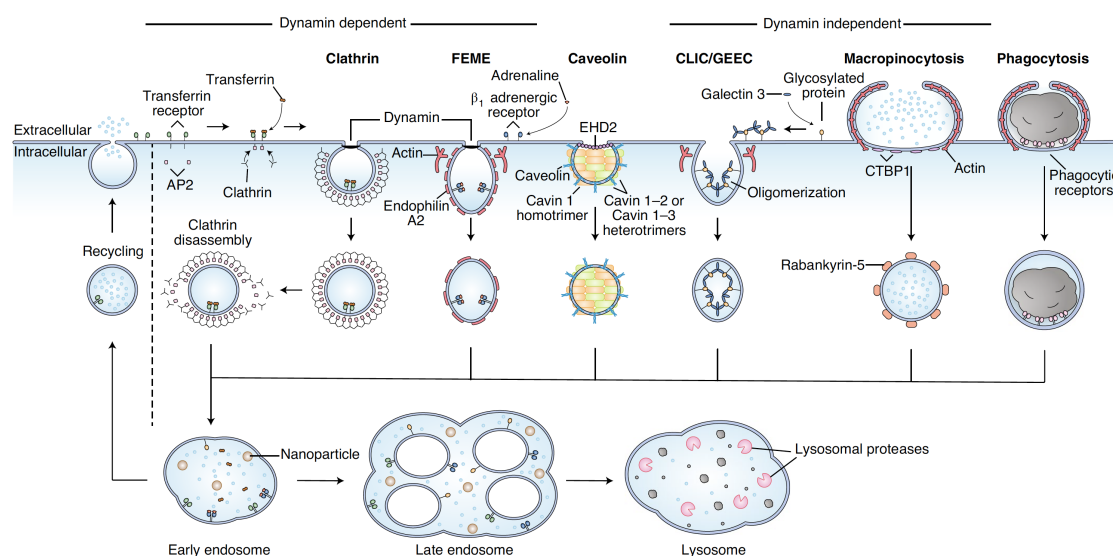
1.4.4 Endolysosomal system

Figure 1.9: *The figure shows the different endocytosis pathways that are frequently discussed in the context of nanoparticle uptake, namely clathrin-mediated endocytosis, fast endophilin-mediated endocytosis (FEME), caveolar endocytosis, clathrin-independent / dynamin-independent endocytosis (CLIC/GEEC), macropinocytosis, and phagocytosis [186]–[188]. The endocytotic carriers formed then fuse with the early endosomes, which sort the cargo and either pass it on to the recycling endosomes or keep it for degradation. The early endosomes mature into late endosomes, which subsequently fuse with the lysosomes. The image was reprinted from [186]. Reproduced with permission from Springer Nature.*

Endocytosis is the process of taking up cargo from the extracellular space through vesicles [189]. Unlike passive uptake mechanisms such as diffusion, endocytosis is a form of active transport into the cell [190]. This process is critical for various functions, including

nutrient uptake, regulation of cellular metabolism, signal transduction, and cell migration [190]–[192]. Endocytosis encompasses various mechanisms, whose discovery and understanding is constantly evolving [191]. As the discussion of endocytosis will be linked to the uptake of CNDs, this text will provide a brief overview of endocytotic mechanisms commonly cited in literature regarding nanoparticle uptake [186]–[188], [193]. The key uptake mechanisms are displayed in Figure 1.9.

The **clathrin-mediated endocytosis (CME)** is the best understood endocytotic pathway and is the main route for the uptake of nutrients [186], [187]. The average diameter of clathrin-coated vesicles is around $100\,nm$, which enables the uptake of nanoparticles up to this size by this route [186], [194]. CME involves the formation of clathrin-coated pits on the cell membrane, which bud inward to form vesicles containing specific cargo molecules [190], [195]. This process comprises of five steps: initiation, cargo selection, coat assembly, scission, and un-coating [195]. CME is initiated by the accumulation of endocytic coat proteins on the inside of the plasma membrane [190]. Additional proteins are recruited from the cytosol to further expand the protein coat [195]. The cargo is recruited by special proteins and concentrated in the coated region [190]. As soon as the coat expands, the membrane bends, resulting in the formation of a clathrin-coated pit [193]. The neck of the pit is then separated from the plasma membrane by the scission process [195]. After separation from the plasma membrane, the clathrin coat is detached from the vesicle, which can now fuse with an endosome [190].

Fast endophilin-mediated endocytosis (FEME) is a pathway that was first described almost 10 years ago [196]. In contrast to CME, FEME is clathrin-independent but dynamin-dependent [186]. Initiated by the activation of specific receptors, endophilin-containing FEME carriers are formed, which detach from the plasma membrane by a scission process [197], [198]. The resulting carriers are tubular with a diameter of $60-80\,nm$ and a length of several hundred nanometres [196]. The attribute "fast" in the name stems from the fact that the entire process is completed just within 4 and 10 seconds [197]. FEME is relevant for growth factor signalling and cell migration [186].

Clathrin-independent/dynamin-independent endocytosis (CLIC/GEEC) is neither clathrin nor dynamin dependent and, similar to FEME, also forms tubular/ring-like carriers with a length of $200-600\,nm$ and a width of $40-80\,nm$ [186], [199]. In contrast to FEME, CLIC/GEEC endocytosis is a constitutive pathway and takes place on a continuous basis in cells that have this pathway [186]. CLIC/GEEC is regarded as a major pathway for the internal uptake of the fluid phase, glycosylated membrane proteins, and lipids [200].

Caveolar endocytosis occurs through invaginations of the plasma membran, resulting in flask-shaped caveolae with a diameter of around $50-80\,nm$ [188]. The proteins caveolin and cavins play important roles in this process and initiate the invagination of the membrane [193]. Caveolae are not found in all cell types (e.g. neurons and many blood cells), but are very common in others [186]. This puts caveolae in contrast to clathrin-coated pits, which have a rather constant density in different cell types [186].

Phagocytosis is primarily associated with immune cells such as macrophages and neutrophils, where large particles ($> 0.5\,\mu m$) such as microbial pathogens, cell debris, or apoptotic cells are engulfed into phagosomes for degradation and antigen presentation

[193]. The process of phagocytosis comprises four steps: i) recognition of the particles to be ingested, ii) activation of the ingestion process, iii) formation of the phagosomes and iv) maturation of the phagosomes [201].

Macropinocytosis involves the non-selective uptake of extracellular fluid and solutes through large ($> 0.5\,\mu m$), actin-driven membrane protrusions called macropinosomes [191], [193]. It serves as a mechanism for nutrient uptake and sensing, signalling, presentation of antigens and cell migration [189]. In contrast to other endocytic pathways, the process is not initiated by cargo molecules binding to the corresponding receptors or particles making contact with the cell surface [189]. Instead, macropinocytosis is induced by the polymerisation of actin on the plasma membrane to form protrusions known as membrane ruffles [189].

After the cargo has been taken up via the pathways, the endocytotic carriers fuse with the early endosomes, which perform a sorting process and pass on materials to be recycled to the recycling endosomes [186], [191]. The early endosomes mature into late endosomes, which finally fuse with the lysosomes to form endolysosomes, where degradation takes place [191], [202]. Lysosomes are discussed in more detail in section 1.4.4.

The uptake of nanoparticles depends on a broad array of factors factors such as size, shape, stiffness and surface properties of the particles. These factors are briefly discussed below.

The **size** of nanoparticles plays a decisive role in their uptake by endocytosis [187]. As described above, the vesicles formed have different sizes. Particles larger than $0.5\,\mu m$ can only be taken up by phagocytosis and/or macropinocytosis. With a size of less than $100\,nm$, however, size is a less important factor in uptake [186]. With regard to the size of the nanoparticles, however, it should be noted that a protein corona is formed when the nanoparticles are added to the cell medium or applied in vivo [186]. The formation of the protein corona can lead to aggregation of the particles, so that the effective size can be considerably larger than measured ex vivo in water or PBS [186].

Just like the size, the **shape** of the nanoparticles also plays an important role. Vácha et al. [203] used simulations to investigate the uptake of nanoparticles of different shapes. They found that the uptake of sphero-cylinders was increased as compared to spheres with the same diameter. Chithrani et al. [204] also found a form-dependent uptake in vitro. The results showed that spherical gold nanoparticles exhibited a five-fold increase in uptake compared to rod-shaped. Possible explanations might include differences in the curvature of the membrane, less available receptor binding sites, influence of surface molecules on the binding of serum proteins to the nanoparticle surface or an uneven protein corona resulting in a lack of multivariate binding to receptors [204].

Another parameter that can influence uptake is the **stiffness** of the nanoparticles. However, the research results are inconclusive; there are studies that show that stiffer nanoparticles are more likely to be absorbed than softer ones and vice versa [205], [206]. While ambiguity exists regarding stiffness, the effects of stiffness appear to depend on the size of the nanoparticles [207]. Overall, stiffness of the nanoparticles can have an influence on uptake and it may be worth including this parameter in the considerations [208].

The **properties of the surface** of the nanoparticles, such as surface charge or surface

modifications, can have a significant influence on the uptake of the particles [187], [208]. Due to the negative charge of the cell membrane, it has a greater affinity for positively charged nanoparticles [209]. However, positively charged nanoparticles can impair the intactness of the cell membrane and therefore increase toxicity [210], [211]. Negatively charged nanoparticles, on the other hand, are taken up more effectively than neutrally charged ones, so charge appears to be a factor in uptake [209], [212]. Modifications to the surface of the nanoparticles with functional groups can have an influence on uptake by changing the charge [208]. On the other hand, modification with so-called targeting ligands such as peptides, small molecules, proteins, antibodies or nucleic acids can force targeted interactions with cell membrane receptors [193], [213].

Lysosomes

Lysosomes are membrane-bound organelles with a size varying between $100\,nm$ and more than $1\,\mu m$, responsible for the degradation and recycling of endocytosed material, cellular waste, and damaged organelles [214], [215]. Previously recognized as pure waste bag of the cell, their critical role in a variety of processes has just been confirmed, such as the maintenance of cellular homeostasis, cell adhesion and migration, gene regulation and metabolic signaling [214], [216], [217]. Lysosomes contain more than 60 acid hydrolases, which degrade the macromolecules that reach the lysosomes via different pathways such as endocytosis, phagocytosis and autophagy [218]. The acid hydrolases function optimally at a pH between 4 and 5 [219]. The lysosomal pH is maintained in this range by proton pumping v-ATPases, chloride channels and ion transporters [219]. Because the lysosomes fulfill many important functions, their malfunction can have serious consequences, ranging from rare lysosomal storage disorders (LSDs) to common autoimmune and neurodegenerative diseases [220]–[223].

1.4.5 Autophagy

Autophagy is a cellular process in which cytoplasmic cargo is transported to the lysosomes for degradation [224]. The main task of autophagy is to supply the cell with nutrients to maintain cellular functions during starvation or other forms of stress [225]. Therefore, autophagy is an adaptive process that is very important for cell homeostasis, both in physiological and pathological situations [226], [227]. Changes in autophagy could be associated with diseases such as cancer, neurodegeneration and cardiac disorders [227]. Transport of the cargo can take place in three different ways, the best researched of which is macroautophagy [228]. Other pathways are microautophagy and chaperone-mediated autophagy [227]. In the following, the focus is on macroautophagy, hereafter referred to as autophagy.

Autophagy can be non-selective or selective [229]. Both involve the formation of autophagosomes, which are vesicles with a double membrane that subsequently fuse with the lysosomes and form the so called autolysosomes in which the degradation of the cargo by the lysosomal hydrolases takes place [230]. In non-selective autophagy, cytoplasm is randomly taken up into the phagophores, the precursors of autophagosomes, in response to stress such as nutritional deficiency [229]. In selective autophagy, the selective uptake of e.g. damaged cell organelles, invading bacteria or aggregated proteins can be triggered by certain signals or cellular events [225]. Various markers can be used to measure the autophagic flux. Among the most common are the protein p62/SQSTM1 (p62) and the microtubule-associated protein 1 light chain 3B (LC3) [228], [231]. LC3 is located in autophagosomes inside and outside the membrane and thus provides information about the

number of autophagosomes and autophagy-related structures [228]. p62 acts as a selective autophagy receptor for the degradation of labeled structures [231]. Proteins p62 and LC3 located within the autolysosome are degraded and are therefore relevant as markers for the autophagic flux [228], [231].

2. Experimental methods

This chapter presents the methods used in the publications resulting from this work. The exact details can be found in the methods section of the publications. First, the method flow cytometry is presented. Afterwards, the synthesis and modification of CNDs are discussed and a brief characterization is given. Confocal fluorescence microscopy is then described. Finally, the methods Western blot and mass spectrometry are introduced.

2.1 Flow cytometry

The method of Flow Cytometry enables the analysis and differentiation of single cells by their optical and fluorescence properties [15]. Cell characteristics like size, granularity, and antigen expression on the cell surface can be determined for each cell [232]. Likewise, it is possible to deduce the number of absorbed fluorescent particles like carbon nanodots. Specialized types of flow cytometers also permit to sort cells based on their fluorescent properties [8]. These flow cytometers are called "fluorescence-activated cell sorter" (FACS) [233].
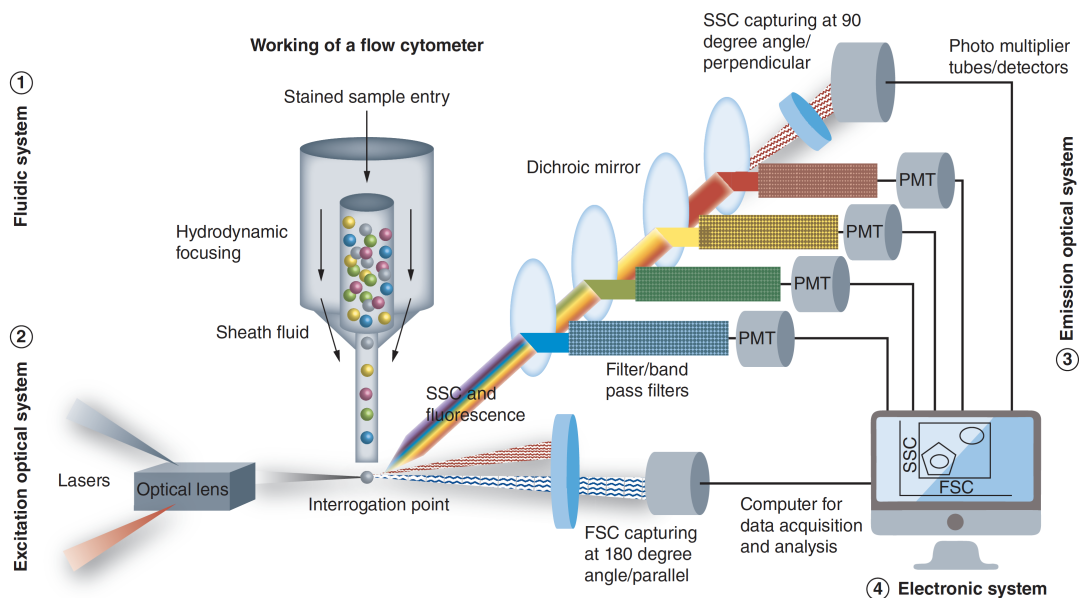
Figure 2.1: *Shown is a schematic illustration of a flow cytometer consisting of the four components: Liquid system, the optical systems (excitation and emission) and the electronic system. The stained cells are separated by hydrodynamic focusing and pass the interrogation point where laser light hits them. The forward-scattered light (FSC) is detected parallel to the excitation, while the side-scattered light (SSC) and the fluorescence signals are detected perpendicular to the excitation. The image was reprinted from [234], ©(2021) reprinted by permission of Informa UK Limited, trading as Taylor & Taylor & Francis Group.*

Figure 2.1 shows a schematic illustration of a flow cytometer. The setup includes four primary components: the fluidic system, the optical systems (excitation and emission), and the electronical system [8]. For the measurement the cells must be present as cell suspension [235]. For the work with adherent cells detachment agends such as trypsin must be used. The sample fluid is injected into the sheath fluid with high air pressure so that the cells align and single cells pass the laser beam [235]. This is called hydrodynamic focusing [8]. The aligned cells then sequentially pass through the laser beam at the interrogation zone. The resulting scattered light and fluorescence signals are split by mirrors, filtered according to wavelength and recorded in the respective detectors [8]. The scattered light varies depending on cell size and granularity. The forward-scattered light (FSC) is associated with cell size and is measured parallel to the excitation beam, while the side-scattered light (SSC) depends on cell granularity and is measured perpendicular to the direction of the laser beam [15]. The intensity of emitted fluorescence light is proportional to the number of bound fluorochrome-labeled antibodies [235]. The optical signals are converted into electronic signals, which can be analyzed.

2.1.1 Gating

As modern flow cytometers incorporate several lasers with different wavelengths and the number of available fluorescence markers continues to increase, multidimensional data sets with up to 30 dimensions emerge accordingly [15], [17]. Due to the increasing number of dimensions, the complexity of subsequent data analysis continues to escalate and becomes more time-consuming [236]. Traditionally, data evaluation is manually performed by a human operator using a method known as gating [8]. In this process, various dimensions (such as FSC, SSC, or fluorescence markers) are plotted against each other in two-dimensional scatterplots, and clusters are defined by manually drawn gates that divide the cells into positive and negative in relation to the respective marker. [8]. These gates assign specific cell populations based on the fluorescence intensities of the plotted markers. Pre-gating is performed beforehand to exclude dead cells, debris, and doublets, ensuring that the subsequent analysis is conducted only on "real" cells [235].
Although traditional analysis is very time-consuming and can vary depending on the operator, manual gating still remains the standard in many clinics [90]. Additionally, two-dimensional projections leads to the loss of high-dimensional relationships [19]. Over the last two decades algorithms have been developed aiming to automate the gating process through clustering and visualize high-dimensional data sets through dimensionality reduction [18], [22], [237].

2.2 Carbon nanoparticles

2.2.1 Synthesis

For the preparation of CNDs used in this thesis, a modified version of the microwave-assisted synthesis bottom up approach described by Qu et al. in [238] was used. This approach was adapted to a microwave reactor by Fasbender et al. [148]. In a first step, $210\,mg$ citric acid and $340\,mg$ diethylenetriamine (DETA) are placed in a microwave reaction vessel and stirred at $400\,rpm$ for 10 minutes. Subsequently, the sealed vessel is placed in an industrial microwave oven manufactured by *CEM* and is heated with continuous stirring to a temperature of $180°C$, which is maintained for $150\,s$. The resulting product is diluted with deionized water (DI water) and is filled into a $10\,ml$ dialysis tube with a

cut-off size of 100-500 Da. The solution is dialyzed against $2\,l$ of DI water. Three water exchanges are performed every 12 hours. After 48 hours, the dialysis is considered complete, and the solution is removed from the dialysis bag. Subsequently, the dialysed product is freeze-dried for further processing like dissolving in phosphate-buffered saline (PBS) or for modifications described below. Freeze-drying was carried out with the kind assistance of the Research Group of Prof. Hartmann, Institute of Macromolecular Chemistry, Heinrich-Heine-University Düsseldorf.

2.2.2 Modifications

To investigate the influence of glycofunctionalization on the selective uptake of CNDs in *Paper III*, these were modified with various monosaccharides and glycooligomers. The coupling was carried out using EDC and NHS to activate the caboxyl groups. In *Paper IV*, the CNDs were modified with bPEI as a test molecule to be delivered by the CNDs as carriers into the lysosomes. EDC and NHS were also used for coupling, the exact details of which can be found in *Paper IV*.

2.2.3 Characterization

(a)   (b)
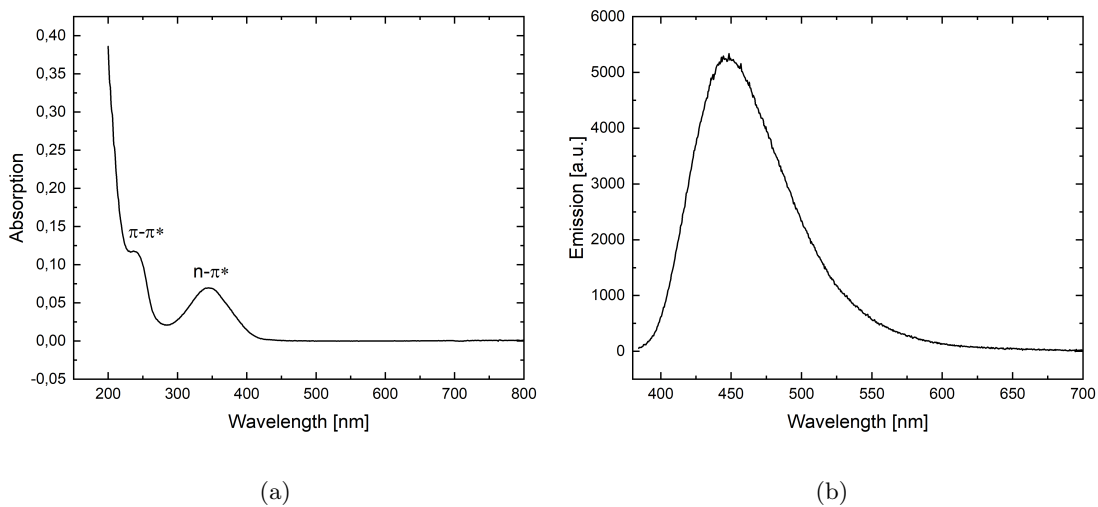
Figure 2.2: *Optical spectra of the CNDs used in the work related to this thesis. (a) Absorption spectrum of the CNDs. The shoulder around 230 nm is attributed to the π-π*-transition, while the peak around 350 nm is attributed to the n-π*-transition according to the literature [238], [239]. b) Emission spectrum of the CNDs, the maximum emission is around 450 nm under exicitation at 360 nm wavelength.*

The composition and structural properties of the CNDs synthesized as described in 2.2.1 were extensively characterized by Stefan Fasbender in studies related to his dissertation. This was done using transmission electron microscopy (TEM), atomic force microscopy (AFM), CHN elemental analysis, X-ray photoelectron spectroscopy (XPS), and Raman spectroscopy [148], [240]. This section provides a brief description of the essential findings. TEM and AFM measurements shed light on the geometry of the CNDs. The particles have an average diameter of $3.3\,nm$ and a hexagonal structure with a lattice constant of $0.223\,nm$. Their height ranges between $1\,nm$ and $2\,nm$ which concurs to two to three layers

of graphene. The CHN elemental analysis showed that the CNDs consist of 40% carbon, 8% hydrogen, and 19% nitrogen (mass fraction). The remaining portion was allocated to oxygen, which aligns with the XPS data. Raman spectroscopy showed the presence of a G-band at $1596\,cm^{-1}$, which corresponds to graphitic sp^2-carbon. Additionally, a D-band was observed at $1375\,cm^{-1}$, which is ascribed to disordered sp^2-carbon. There were also Dl-, D2-, and D3-bands at 1195, 1264, and $1412\,cm^{-1}$, which may be assigned to various sp^3-carbon containing groups [241].

Figure 2.2 displays an absorption and emission spectrum of the CNDs. The absorption spectrum of the nanoparticles shows a shoulder around $230\,nm$ and a peak around $350\,nm$, which can be attributed to a π-π*-transition respective a n-π*-transition according to the literature [238], [239]. The maximum emission occurs at approximately $450\,nm$ when excited at around $360\,nm$. The quantum yield, as determined by the method of Williams et al. [242], is approximately 25%.

2.3 Confocal fluorescence microscopy

The technique of confocal fluorescence microscopy is well known. Since it forms the basis for investigating the intracellular distribution of CNDs in this thesis, it will be explained briefly. This section is based primarily on the references [243] and [244].

Confocal fluorescence microscopy, also known as confocal laser scanning microscopy, utilizes pinholes in both the illumination and detection planes to ensure that only focused light reaches the detector. The concept of confocal scanning microscopy was invented by M. Minsky in 1955. Later the technique was further developed by M.D. Egger and P. Davidovits, who constructed the first confocal laser scanning microscope in 1969. Figure 2.3 illustrates the setup.
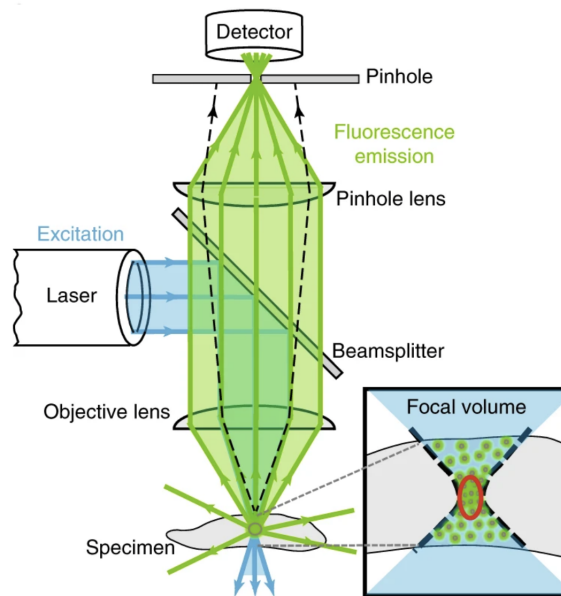

Figure 2.3: *Shown is a schematic representation of the main components of a confocal fluorescence microscope and the beam path. The light from the excitation beam (blue) is focused on the sample. The fluorescence light from the focal volume (green) is focused by lenses in such a way that it passes through the pinhole and can be captured by the detector. Light from outside the focus (black dashed line) is blocked by the pinhole. The image was reprinted from [245]. Reproduced with permission from Springer Nature.*

This technique utilises a dichroic mirror or beam splitter to collimate and direct laser light onto the objective lens. The objective lens then focuses the excitation beam onto the sample, exciting the fluorophores within it. The emitted fluorescence light is then focused on a conjugate plane from the tube lens, where a pinhole aperture selectively allows light from the confocal plane to reach the detector, while blocking out-of-focus light. Typically, photo-multiplier tubes (PMTs), avalanche photodiodes (APDs), or charge-coupled device (CCD) cameras are used as detectors in this system. As the illumination pinhole selectively illuminates the point of interest, this necessitates sequential scanning in the x and y directions to obtain the intensity for each sample point to create a 2D image. This technique offers several advantages: reducing background light from unfocused layers, enhancing spatial resolution, and enabling imaging of 3D samples when combined with z scanning. By exclusively detecting in-focus light, confocal microscopy can produce images of thin sample sections.

2.4 Western Blot

The Western Blot is a method used to separate and identify proteins. It was developed by Towbin et al. in 1979 [246]. This technique is used in *Paper IV* to determine the concentrations of the proteins p62 and LC3, which serve as biomarkers of autophagy [228], [231]. The aim is to investigate the influence of CNDs on autophagy and on cellular metabolism. The experimental details are given in *Paper IV*, the following section provides a general overview of the method, building upon two references [247], [248].
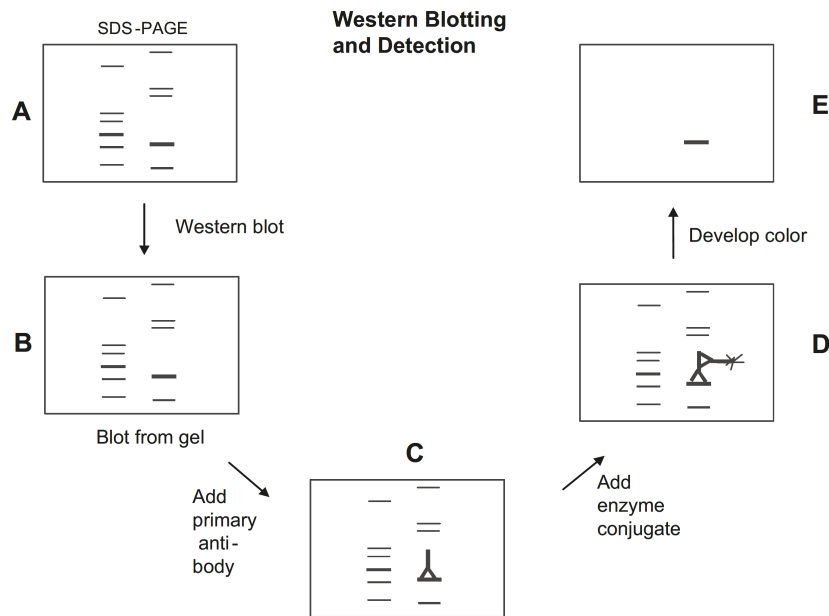

Figure 2.4: *The figure shows schematically the different steps of the Western blot. a) Unstained SDS-PAGE gel before the Western blot. b) Replicate of the gel on the membrane after the transfer. c) Staining with the primary antibody. d) Staining with the secondary antibody. e) Specific protein band visible by fluorescence signal upon excitation. The image was reprinted from [249], ©(2006) with permission from Elsevier.*

Figure 2.4 depicts the schematic procedure of a Western Blot. In a first step the cells are lysed and the proteins are extracted from the cell lysate. Afterwards the protein

concentrations are determined. Sodium Dodecyl Sulfate (SDS) is used to mask the charges of the proteins after denaturation, resulting in an overall negative surface charge of the proteins. The proteins are then separated by molecular weight through gel electrophoresis, by loading the proteins onto a polyacrylamide gel and applying an electric field. Due to the negative surface charge the proteins migrate through the gel towards the positive electrode. The gel acts as a molecular sieve, allowing smaller proteins to move faster towards the anode than larger proteins [250]. This results in the formation of bands of proteins of similar size within the gel after a defined period of time. In the blotting step, the proteins are transferred onto a membrane, fixed in specific locations, allowing their storage for further experiments. The transfer can be done using different methods. In electroblotting, which is the most commonly used method, the proteins are transferred from the gel to the membrane using an electric field applied perpendicular to the gel. This step is also referred to as immunoblotting. After the transfer the membrane is blocked in regions where no proteins adhere. Following this, the proteins are labeled and stained with specific antibodies and fluorophores in two steps. First the membrane is incubated with a solution of primary antibody for the specific protein. In a second step, the membrane is exposed to the secondary antibody which is specific for the host spicies of the first antibody and is labeled with fluorophores allowing optical detection. Since several secondary antibodies may bind to one primary antibody the fluorescence signal is enhanced. This enables a comparison of protein amounts based on fluorescence intensity between different samples.

2.5 Mass spectrometry

This section is primarily based on the references [251] and [252].

Mass spectrometry (MS) is an analytical method for determining the molecular mass of molecules, peptides and proteins. While the Western blot method described above can be used to detect individual proteins, MS can be used to analyze the entire proteome of a cell. A schematic representation is shown in Figure 2.5.
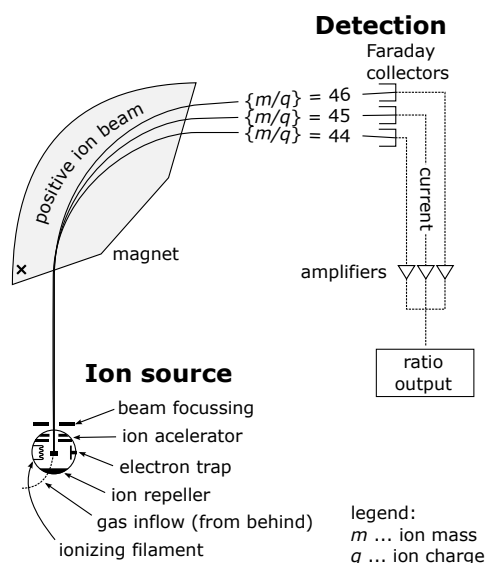

Figure 2.5: *Schematic representation of a mass spectrometer. It essentially consists of three components, the ion source, the mass analyzer and a detector. By analyzing the mass-to-charge ratios, atoms, molecules and proteins can be determined. The image is public domain.*

The MS method was used in *Paper IV* to analyze the proteome and secretome. The analysis was carried out by the MPL (Molecular Proteomics Laboratory) at the BMFZ (Biologisch-Medizinisches Forschungszentrum) of the HHU Düsseldorf by Thomas Lenz. The exact experimental details are laid out in the methods section of *Paper IV*.

For the analysis, the molecules (here: peptide fragments after protein digestion) are transferred into the gas phase and ionized using an ion source. An applied electric field accelerates the ions, which are separated in the mass analyzer according to their mass-to-charge ratio $\frac{m}{q}$. There are different types of mass analyzers and all are based on the use of static and/or variable electric and magnetic fields. The whole process takes place under high vacuum conditions. The separated ions are then recorded by a detector. The protein abundances can be reconstructed from the detected peptide sequences afterwards.

3. AI assisted analysis of multiparameter flow cytometry data

As described in section 1.2.1 the number of parameters that can be measured simultaneously in flow cytometry is constantly increasing, which makes the analysis increasingly complex and time-consuming. AI-based methods seem promising to face these challenges. *Paper I* explores the question how the t-SNE algorithm can contribute to the analysis of hematopoietic stem and progenitor cells (HSPCs) and support the diagnosis of the diseases AML and MDS. The immunological phenotype of HSPCs were examined based on the differential expression of various markers using a created t-SNE map and compared them with the immunophenotype of blasts and leukemic stem cells (LSCs) in AML and MDS. It was shown that the defined clusters on the t-SNE map could be assigned to diagnostically relevant cell populations and allowed discrimination at a detailed level. Furthermore, a method was presented for classifying new samples using t-SNE reference maps based on a similarity comparison with the pearson coefficient as a quantitative measure.

As an additional parameter in the t-SNE assisted evaluation, the uptake of carbon nanoparticles was examined to differentiate between hematopoietic cells from healthy donors and from patients with AML in *Paper II* . There were differences in the uptake between the two groups, which could be visualized clearly using t-SNE. While all cell types of the normal and leukemic cells took up CNDs, the $CD34^+$ and $CD33^+$ subsets of the AML samples showed a significantly reduced uptake compared to the subsets of the healthy donors. In addition, it was observed that the CNDs accumulate in lysosomes of leukemic blasts. These observations in *Paper II* gave rise to new questions that are addressed in *Paper III* and *Paper IV* in Chapter 4.

3.1 Paper I

Reference

Reproduced from [C. Nollmann, W. Moskorz, C. Wimmenauer, P.S. Jäger, R.P. Cadeddu, J. Timm, T. Heinzel and R. Haas (2024). Characterization of CD34$^+$ Cells from Patients with Acute Myeloid Leukemia (AML) and Myelodysplastic Syndromes (MDS) Using a t-Distributed Stochastic Neighbor Embedding (t-SNE) Protocol. Cancers, 16(7), 1320. https://doi.org/10.3390/ cancers16071320], with the permission of MDPI.

Copyright statement

Contributions

I participated in the conception and development of the method. I developed the code together with CW. I participated in the data collection, curation, analysis and verification. I wrote the initial draft of the paper together with TH and RH and prepared the figures.

Characterization of CD34+ Cells from Patients with Acute Myeloid Leukemia (AML) and Myelodysplastic Syndromes (MDS) Using a t-Distributed Stochastic Neighbor Embedding (t-SNE) Protocol

Cathrin Nollmann [1], Wiebke Moskorz [2], Christian Wimmenauer [1], Paul S. Jäger [3], Ron P. Cadeddu [3], Jörg Timm [2], Thomas Heinzel [1,*] and Rainer Haas [3,*]

[1] Condensed Matter Physics Laboratory, Heinrich-Heine-University, 40204 Düsseldorf, Germany; cathrin.nollmann@hhu.de (C.N.)
[2] Institute of Virology, Heinrich-Heine-University, 40204 Düsseldorf, Germany; joerg.timm@med.uni-duesseldorf.de (J.T.)
[3] Department of Hematology, Oncology and Clinical Immunology, Medical Faculty, Heinrich-Heine-University, 40225 Düsseldorf, Germany; paulsebastian.jaeger@med.uni-duesseldorf.de (P.S.J.)
[*] Correspondence: thomas.heinzel@hhu.de (T.H.); haas@med.uni-duesseldorf.de (R.H.)

Simple Summary: Hematopoietic stem and progenitor cells (HSPCs) play a pivotal role in maintaining the homeostasis of the blood and immune systems. Acute myeloid leukemia (AML) and myelodysplastic syndromes (MDS) represent heterogeneous hematologic malignancies resulting from genetic mutations within cells of the hematopoietic lineage, leading to the expansion of leukemic blasts including leukemic stem cells (LSCs). Using the t-distributed stochastic neighbor embedding (t-SNE) methodology, we examined the immunological phenotype of HSPCs based on the differential expression of CD34, CD38, CD45RA, CD123 and programmed death ligand 1 (PD-L1) antigens, and contrasted it with the immunophenotype of blasts and LSCs in AML and MDS.

Abstract: Using multi-color flow cytometry analysis, we studied the immunophenotypical differences between leukemic cells from patients with AML/MDS and hematopoietic stem and progenitor cells (HSPCs) from patients in complete remission (CR) following their successful treatment. The panel of markers included CD34, CD38, CD45RA, CD123 as representatives for a hierarchical hematopoietic stem and progenitor cell (HSPC) classification as well as programmed death ligand 1 (PD-L1). Rather than restricting the evaluation on a 2- or 3-dimensional analysis, we applied a t-distributed stochastic neighbor embedding (t-SNE) approach to obtain deeper insight and segregation between leukemic cells and normal HPSCs. For that purpose, we created a t-SNE map, which resulted in the visualization of 27 cell clusters based on their similarity concerning the composition and intensity of antigen expression. Two of these clusters were "leukemia-related" containing a great proportion of CD34+/CD38− hematopoietic stem cells (HSCs) or CD34+ cells with a strong co-expression of CD45RA/CD123, respectively. CD34+ cells within the latter cluster were also highly positive for PD-L1 reflecting their immunosuppressive capacity. Beyond this proof of principle study, the inclusion of additional markers will be helpful to refine the differentiation between normal HSPCs and leukemic cells, particularly in the context of minimal disease detection and antigen-targeted therapeutic interventions. Furthermore, we suggest a protocol for the assignment of new cell ensembles in quantitative terms, via a numerical value, the Pearson coefficient, based on a similarity comparison of the t-SNE pattern with a reference.

Keywords: hematopoietic stem and progenitor cell (HSPC); acute myeloid leukemia (AML); myelodysplastic syndromes (MDS); leukemic stem cells (LSC); CD34; CD38; CD45RA; CD123; PD-L1; flow cytometry; t-SNE; high-dimensional space analyses; classification; dimensionality reduction; immunophenotyping

1. Introduction

Acute myeloid leukemia (AML) and myelodysplastic syndromes (MDS) are heterogeneous disorders originating from hematopoietic stem cells (HSCs) through the progressive and sequential acquisition of genetic and epigenetic alterations. As a result, there is a clonal expansion of myeloid progenitors/precursors in the bone marrow (BM) and peripheral blood (PB), associated with impaired cell differentiation leading to hematopoietic insufficiency [1–3]. Drug resistance and dormancy of the leukemic stem cells (LSC) with a reduced susceptibility to cytotoxic drugs are the main reasons for treatment failure [4,5]. To tackle the problem of resistance and dormancy, a subtle characterization of the leukemic blast population including the LSC is a prerequisite for a more efficacious targeting and eradication.

The search for a better characterization of the various subsets contained within the bulk mass of leukemic blasts from patients with acute leukemia has prompted a constant increase in the number of antigens in panels for single-cell cytometry, reaching numbers from 14 to 28 colors [6–8]. In order to visualize and interpret that kind of multidimensional marker expression function (MEF), the traditional representation via two-dimensional scatter plots increases correspondingly and reaches its limits.

Mathematically, these scatter plots represent two-dimensional projections of the multidimensional function, by which some information of the original distribution is inevitably lost. Consequently, rare or so far unknown leukemic subpopulations of pathophysiological relevance may be not detected [9,10] or multi-dimensional structural information may get lost. This prompted the search for two-dimensional, graphical representations of the MEF, which preserve the full representations [11]. Among such mapping algorithms, t-distributed stochastic neighbor embedding (t-SNE) [12] is a promising candidate [13–17]. We opted for the t-SNE algorithm as compared to equally valid alternatives like uniform manifold approximation and projection (UMAP) [18] since it has a long and successful track record over the past decade and is one of the most widely used for comparable tasks [19,20]. Using this methodology, we aimed at elucidating the CD34$^+$ cell population in more depth in samples from patients with AML and MDS in comparison to samples from patients in complete remission (CR) following antineoplastic therapy.

2. Materials and Methods

2.1. Patients

BM samples of 21 patients with MDS (6 patients) and AML (15 patients) were obtained at the Department of Hematology, Oncology and Clinical Immunology from the University Hospital Düsseldorf on their regular follow-up visits for routine diagnostics. Our control population consisted of 12 patients following allografting, cytotoxic chemotherapy or both who were in CR, with 2 patients (#8 and #21 marked with asterisks in Table 1) still not having achieved full hematological reconstitution. The characteristics of the entire group of patients are shown in Table 1.

Table 1. Patient characteristics. The samples are grouped according to whether the patients have active disease (AD) or are in complete remission (CR). All patients in CR are MRD negative (see Supplemental Table S2 for the MRD analysis).

Group	Pat. ID	Age	Sex	WHO Classification	Status of Disease	Initial Mutation	Cytogenetic	Time ** (Months)
AD	1	61	m	MDS-IB2	AD	-	46, XY	11
AD	4	58	f	AML	Mr	-	47, XX, +11	10
AD	10	58	m	MDS-IB2	Hr	-	46, XY	30
AD	11	67	m	AML, md-r	Mr/p	*FLT3-ITD, RUNX1, EZH2*	46, XY	5
AD	12	65	f	AML, md-r	Hr	*ASXL1*	46, XX, del(11)(q21,q24) [21]	26

Table 1. *Cont.*

Group	Pat. ID	Age	Sex	WHO Classification	Status of Disease	Initial Mutation	Cytogenetic	Time ** (Months)
AD	13	60	m	AML	Id	*IDH1*	47, XY, +8[22]/46, XY[2]	0
AD	14	69	f	AML, md-r	Hr	*JAK2*	45, XX, -7	39
AD	16	60	m	AML, md-r	Hr	*ASXL1, RUNX1*	not initial: 46, XY, del(3)(q21q25)[23]/47idem+8[5]	53
AD	17	60	f	AML with minimal differentiation	Hr	*IDH2*	47, XX, +mar[4]/46, XX [22], cytogenetic aberration: 7(4;12)	216
CR	2	76	m	MDS-IB2	CR	*ASXL1*	46, XY	13
CR	3	67	f	AML with CEBPA mutation	CR	*CEBPA*	46, XX	21
CR	5	56	f	AML with maturation	CR	*DNMT3A, IDH1*	46, XX	4
CR	6	41	m	AML, md-r	CR	*RUNX1*	complex karyotype	3
CR	7	54	f	AML with NPM1 mutation	CR	*NPM1, IDH2*	46, XX	8
CR	8	51	m	MDS with low blasts and SF3B1 mutation (MDS-SF3B1)	CR *	*JAK2, SF3B1*	complex karyotype	60
CR	9	67	f	AML with CBFB-MYH11 fusion	CR	*CBFB-MYH11*	46, XX, inv(16)(p13q22)[24]/ 46, XX [3]	29
CR	15	28	f	AML, md-r	CR	*RUNX1*	complex karyotype	47
CR	18	39	f	AML, md-r	CR	*FLT3-ITD*	del(7)(q22[22]/46, XX [3]	2
CR	19	40	m	AML, md-r	CR	*ASXL1, c-KIT, TET2*	+8, XXY, add(21p)	32
CR	20	61	f	AML, md-r	CR	*ASXL1, RUNX1*	46, XX	57
CR	21	70	m	AML, md-r	CR *	*ASXL1, RUNX1, TET2, EZH2*	46, XY	11

* Patients who still not having achieved full hematological reconstitution; ** time difference between initial diagnosis and sample collection; md-r: myelodysplasia-related; CR: complete remission; Hr: hematological recurrence; Mr: molecular recurrence; Mr/p: molecular recurrence/persistence; Id: initial diagnosis; AD: active disease; MDS: myelodysplastic syndrome.

2.2. Isolation and Phenotyping of White Blood Cells

White blood cells (WBCs) were isolated via red blood cell lysis. For that, BM was collected in EDTA coated syringes or blood collection tubes and bone fragments were removed by filtering the BM with a 70 μm cell strainer. The BM was then incubated 1:10 for 10 min with isotonic ammonium chloride solution (155 mM NH_4Cl, 10 mM $KHCO_3$ and 0.1 mM EDTA, pH 7.4, purchased from the University Hospital Düsseldorf Pharmacy, Düsseldorf, Germany). WBCs were pelleted for 5 min at 500 g, supernatant was discarded and the remaining WBCs were washed twice with DPBS prior to staining. For each sample, one to four million cells were transferred to a 96 well U-Bottom plate and dead cells were stained with fixable viability dye (#65-08666-14, Thermo Fisher Scientific, 1:1000 in DPBS), washed with DPBS and subsequently stained for cell surface molecules. Dead cells and surface molecules were each stained for 15 min at room temperature in the dark. Antibodies for cell surface molecules (see Table 2) were diluted in Brilliant Stain Buffer (#566349, BD Horizon, BD Bioscience, Franklin Lakes, NJ, USA) to prevent staining artifacts due to polymer dyes. Prior to data acquisition at a BD LSR Fortessa (V/B/YG/R), cells were washed with DPBS, fixed overnight (IC Fixation Buffer, #00-8222, Thermo Fisher Scientific, Waltham, MA, USA) and washed again. Cells were taken up in DPBS and acquired at up to 3000 events/s. All samples contain more than 10^5 cells.

Table 2. WBC staining panel.

Specificity	Clone	Fluorescence Dye	Vendor	Cat #	RRID	Concentration
Fixable viability dye	/	eFlour506	TFS *	65-0866-14	/	1:1000
PD-L1	MIH5	PerCP-eFlour710	TFS	46-5983-42	AB_11041815	1:50
CD123	6H6	PE	TFS	12-1239-42	AB_10609206	1:100
CD45	HI30	PE-Cy5	BioLegend	304010	AB_314398	1:200
CD45RA	HI100	PE-Cy7	TFS	25-0458-42	AB_1548774	1:200
CD34	4H11	APC	TFS	17-0349-41	AB_2016604	1:50
CD38	HIT2	APC-eFlour780	TFS	47-0389-41	AB_11217871	1:50

* TFS: Thermo Fisher Scientific.

2.3. Gating Strategy

In the study presented here, we were particularly interested in a detailed, multi-color flow cytometry-based characterization of the $CD34^+$ cells focusing on a subtle comparison between the $CD34^+$ cell subsets of patients with active disease (AD) and those of patients in CR. For that purpose, we used a panel of the following monoclonal antibodies: CD34, CD38, CD45, CD45RA and CD123, as it provides the basis for defining the various types of HSPCs (Table 3). In addition, the programmed death ligand 1 (PD-L1) was included, since it is also expressed on normal hematopoietic cells, exerting a suppressive effect on the immunological response. We consider this panel suitable and sufficient to demonstrate strengths and pitfalls of a t-SNE-based analysis.

Table 3. Antigen combinations for HSPC characterization.

Cell Type	Label	Antigen Combination
Hematopoietic stem cells	HSC	$CD34^+$ $CD38^-$ ($CD90^+$ not included)
Multipotent progenitor cells	MPP	$CD34^+$ $CD38^-$ ($CD90^-$ not included)
Common lymphoid progenitors	CLP	$CD34^+$ $CD38^-$ $CD45RA^+$
Common myeloid progenitors	CMP	$CD34^+$ $CD38^+$ $CD45RA^-$ $CD123^{low}$ *
Megakaryocyte/erythroid progenitors	MEP	$CD34^+$ $CD38^+$ $CD45RA^-$ $CD123^-$
Granulocyte-macrophage progenitors	GMP	$CD34^+$ $CD38^+$ $CD45RA^+$ $CD123^+$
Not identified by this set of antigens	Other	Various combinations

* By the term "low" we refer to weakly positive.

Our gating strategy for the cells of interest, i.e., the $CD34^+$ cells, encompassed six steps including: (1) an FSC vs. SSC gate, (2) a CD45 vs. SSC gate, (3) an exclusion step for the elimination of doublets, and (4) a viability check using eF506 dye for the exclusion of dead cells. As a result, the (5) final gate of interest (GOI) contained $CD34^+$ cells excluding the population of granulocytes. Afterwards (6), only the $CD34^+$ cells were selected, as shown in Figure 1A and B for patient 1. The gating was carried out in FlowJo® (FlowJo, Ashland, OR, USA). The number of $CD34^+$ positive cells of an individual patient varied between 129 and 207,994 (Figure 1C). To avoid domination of features in the t-SNE plots by individual patients, a maximum of 1000 cells was randomly selected in patients with high cell numbers.
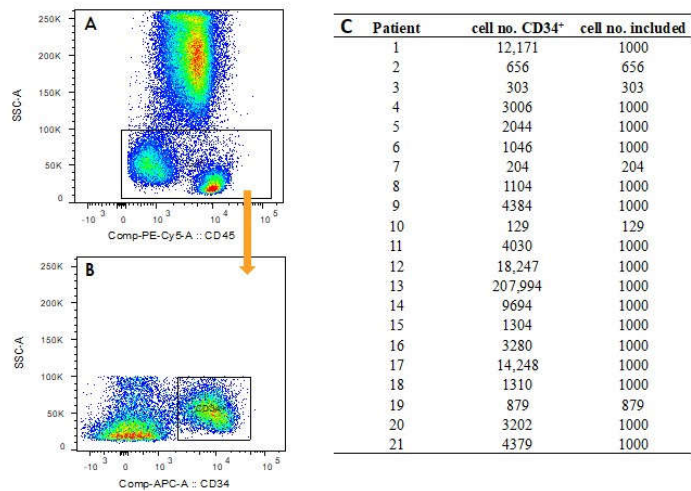
C	Patient	cell no. CD34+	cell no. included
	1	12,171	1000
	2	656	656
	3	303	303
	4	3006	1000
	5	2044	1000
	6	1046	1000
	7	204	204
	8	1104	1000
	9	4384	1000
	10	129	129
	11	4030	1000
	12	18,247	1000
	13	207,994	1000
	14	9694	1000
	15	1304	1000
	16	3280	1000
	17	14,248	1000
	18	1310	1000
	19	879	879
	20	3202	1000
	21	4379	1000

Figure 1. In (**A**) and (**B**), steps (5) and (6) of the gating strategy are shown. The orange arrow means that in B only the cells from the gate of interest (GOI, black frame) were analyzed. (**A**) The exclusion of the population of granulocytes with (5) a GOI; (**B**) The final step (6), selection of the CD34+ cells; In (**C**), the numbers of CD34+ cells after the gating are shown for each patient as well as the numbers thereof included in the t-SNE analysis.

2.4. Visualization by t-SNE

In a second step, we applied t-SNE for the visualization of different cell clusters based on their similarity with regards to the composition and intensity of antigen expression. The t-SNE algorithm is a nonlinear dimensionality reduction technique which visualizes high-dimensional data in a two-dimensional scatter plot in such a way that the clustering in high dimensions is preserved. Cells exhibiting comparable protein-expression patterns are positioned adjacently on the t-SNE map, facilitating the depiction of distinct cellular subgroups. The nature of this algorithm is to preserve the local relationships and not the global structure [21]. This is one of the well-known limitations of t-SNE [22–24]. Thus, global structures such as the arrangement of the clusters and their distances in the t-SNE plane provide no basis for interpretation. A principal component analysis (PCA) was therefore carried out for the initialization to improve the global structure of the plot, as established in the literature [21,25].

To enable comparability between the t-SNE plots, the gated CD34+ cells from all patients were first merged into a common data set. The patient ID and the group assignment were appended to the expression matrix prior to data merger permitting the subsequent separation according to these characteristics after the t-SNE analysis. The fluorescence data were scaled biexponentially in a preliminary step [26]. The t-SNE analysis was carried out using the Barnes–Hut implementation of t-SNE by the Rtsne package (Version 0.16, Open Source). The code is available in the Supplement (Code S1). For the PCA the predefined value of 50 for the number of retained dimensions was used. The perplexity as well as the number of iterations were varied over wide intervals. These variations, shown in Supplement Figure S1, not only reassure us that the structures to be interpreted are robust, but also demonstrate that a perplexity of 70 and 3000 iterations is a reasonable choice providing visibility of the relevant morphology within an acceptable computation time [27]. The t-SNE coordinates (t-SNE1, t-SNE2) were also appended to the expression matrix as novel parameters. The entire data set as well as a data set of only patients in CR and a data set of the patients with AD were then exported as FSC files for further analysis in FlowJo ®. The t-SNE plot is created as a function of the two parameters (t-SNE1 and t-SNE2). Since distances within a t-SNE plot cannot be interpreted in a straightforward way for reasons mentioned above, axis labels are omitted for all t-SNE plots, in agreement with common practice.

2.5. Defining Gates in the t-SNE Plots

The expression matrix with the t-SNE coordinates of the three data sets (All patients, only CR, only AD) were imported into FlowJo®. A group was created therein containing all three data sets. Density-based polygon gates were manually drawn on the common t-SNE plot of all patients in CR. Afterward, the 27 gates were applied to the FlowJo® group in order to transfer them to the remaining two data sets.

2.6. Determination of the Immunological Phenotypes of HSPCs

The immunological phenotype of HSPCs was determined using the markers CD34, CD38, CD123 and CD45RA. For this purpose, the limits for the classification into positive (+) and negative (−) according to the marker expression were determined in FlowJo® on the common data set of all patients using scatter plots (Supplement Figure S2). The fluorescence values for the markers used were exported separately for the patients in CR and patients with AD for all gates and then displayed as boxplots (Supplement Figure S3). The commonly used limits for the classification were drawn into the boxplots with the antigen expression levels and the immunological phenotype of the HSPCs was then determined for the respective gate depending on whether the mean value of the respective marker was above or below the limit.

2.7. Quantitative Comparison of t-SNE Plots Using the Pearson Correlation Coefficient

For the quantitative comparison of the t-SNE plots, the density matrix for the respective t-SNE plot was first calculated in R, e.g., for a single patient or for the cumulative image of all patients with AD. The density matrices were exported and the Pearson coefficients between the t-SNE plots were determined using Python. The code is available in the Supplement (Code S2). The density plots of all patients are shown in Supplement Figure S4.

3. Results and Discussion

3.1. Design of a t-SNE-Based Protocol for Multicolor Flow Cytometry Analysis

For the t-SNE analysis, a common data set of all FSC files from all patients was created, so that the t-SNE plots are comparable between the patients. The t-SNE analysis was carried out based on the expressions of CD34, CD38, CD45RA, CD123 and PD-L1. In general, the CD34 antigen permits the identification of hematopoietic stem and progenitor cells, while CD38 is considered a marker associated with differentiation [28,29]. The combination of these two antigens with CD45RA and CD123 permits a characterization and quantification within a BM of the HSC/HPC within a BM sample [30,31].

After the t-SNE run on the combined data set (Figure 2A), the contributions of the CR and AD patients were visualized separately to recognize their contributions to the combined t-SNE picture (Figure 2B,C).

By t-SNE, the cells are arranged in five islands (I–V) of different sizes. It becomes immediately apparent that the CR (in 2B) and the AD (in 2C) samples contribute almost complementarily to the combined representation (A). While in the CR patients, most of the cells are in the east part of the main island (I) plus in three of the four separated islands (II, III, V), the cells from the AD patients accumulate more to the west of the main island as well as in island IV. However, the populations are not mutually exclusive, as all cell types are present in both groups, albeit in some regions with strikingly different prevalence. This phenomenon is most likely not related to "contaminating" leukemic cells within the CR samples as the CR patients are MDR negative. In the t-SNE representation, the cells are distributed according to marker-specific gradients, as shown in the bottom row of Figure 2 D–H for the combined data set, where the black horizontal bars in the color scale column define the corresponding intensity intervals. The overlay of the different markers is shown separately for AD and CR in Supplement Figure S5, as well as an example for two patients from each of the groups in Supplement Figure S6.
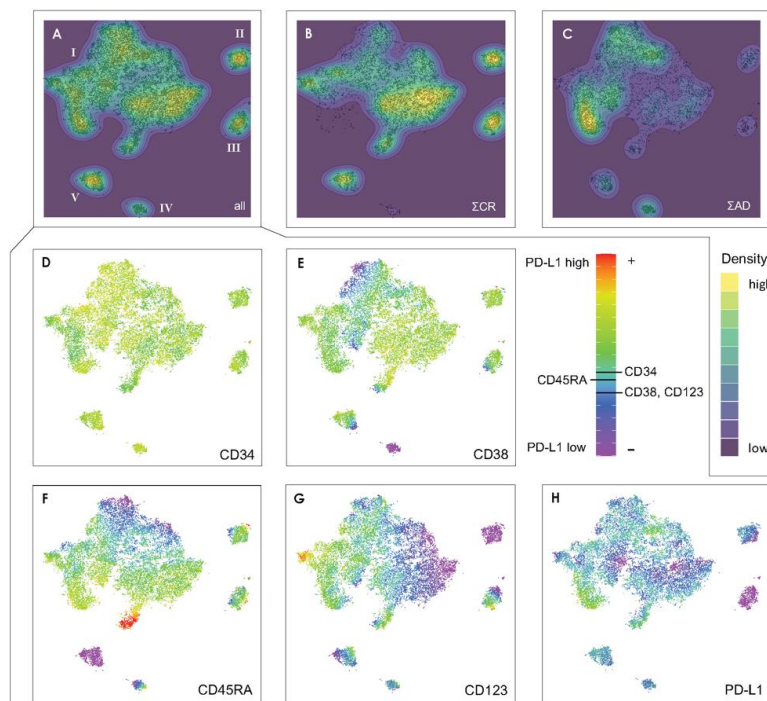
Figure 2. (**A**) t-SNE representation of the combined data set (cells of patients with active disease (AD) and in complete remission (CR)), and the contributions from patients in CR (**B**) as well as from the patients with AD (**C**). The color scale for (**A**–**C**) corresponds to the local density of cells in arbitrary units. In each t-SNE plot, the color scale starts at zero and is normalized to the maximum density in the respective plot. In (**D**–**H**), the prevalence of the various markers entering the t-SNE algorithm are reproduced, for CD34, CD38, CD45RA, CD123 and PD-L1. The color scale represents the expression level. The classification in terms of *positive* (+) and *negative* (−) expression is indicated by the black horizontal lines in the color bar.

As far as CD34 is concerned, the corresponding intensities in (Figure 2D) comprise only positive values since per definition only cells above the threshold of expression were included. Still, the islands in the t-SNE plot show quite varying CD34 expression levels. With respect to CD38, the cells are assorted from northwest to southeast of the main island with increasing expression level, while it is particularly low in island IV. A pronounced CD38 gradient is visible in island III and V, indicating a sub-ensemble of cells undergoing some kind of development. The CD123 concentration, on the other hand, increases from east to west across island I, is almost zero in island II, and shows gradients within islands III and V. The CD45RA expression increases strongly from north to south. Finally, the PD-L1 expression, which is not considered in the assignment of the cells according to Table 3, is non-monotonously distributed across island I and takes characteristic low values in islands II and III.

We can therefore conclude that the gradients in the intensities of CD38, CD45RA and CD123 cause the main substructure in island I, while the expression levels of CD34 and PD-L1 refine this landscape.

3.2. Exemplifying Discussion of t-SNE Gates

For a more detailed study of the five islands, we have defined 27 gates in the t-SNE plot of the CR samples, each with a characteristic set of expression levels for the markers used (Figure 3A). This gate pattern was then transferred without modifications to the AD data as described in Section 2.5, shown in Figure 3B. The percentage distribution of the

cells in the 27 gates for the three datasets (all patients, only CR, only AD) is shown in Supplement Table S1.
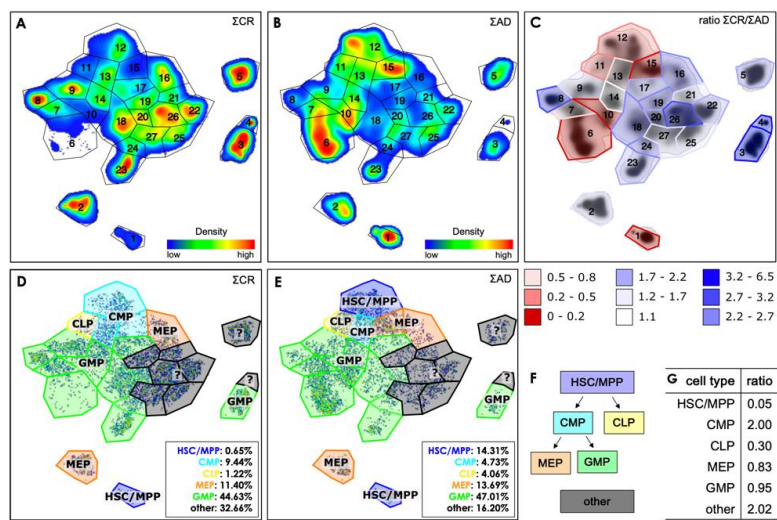
Figure 3. (**A**) Density-based gate definition on the t-SNE plot of the CR data set, performed by visual inspection; (**B**) Application of the gates on the AD data set; In (**C**), the ratio of the percentage distribution of the cells for the CR to the AD samples is given for each gate; The gates are inked in red if this ratio is smaller than 1, i.e., most of the cells in this gate come from AD samples, and blue for ratios larger than 1. The cell type identification for the gates is represented in (**D**,**E**) for the CR samples and for the AD samples, respectively; (**F**) Evolution scheme for the relevant cell types; The ratios of the percentage distribution of cells for the different types in CR samples vs. AD samples are listed in (**G**).

From that kind of visualization, eight gates emerge, namely gates 1, 6, 7, 10–13 and 15, in which the cells of patients with AD dominate. The remaining gates contain more cells from the CR patients, while within gate 14 the ratio is very close to 1.

The box plots of the 27 gates (Supplement Figure S3) were used to assign the cells within each gate according to the classification scheme as detailed in Table 3 and shown in Figure 3F. As can be extracted from Figure 3, the CR subsets are composed of 0.7% HSC/MPP, 9.4% CMP, 1.2% CLP, 11.4% MEP and 44.6% GMP. A proportion of cells (32.7%) could not be allocated according to the classification scheme. On the other hand, the samples of the patients with AD comprised 14.3% HSC/MPP, 4.7% CMP, 4.1% CLP, 13.7% MEP and 47.0% GMP with a proportion of 16.2% of the cells which could not be classified. Clearly, in comparison with the CR samples the AD samples show a significantly greater proportion of HSC/MPP as well as CLP cells, while there are smaller percentages of CMP cells as well as of those cells that cannot be allocated. The fractions of MEP cells are approximately equal for both groups.

Beyond this canonical classification, the t-SNE representation provides a rich substructure within the regions of particular cellular subtypes, reflecting subtle differences between the various populations. A complete delineation of all 27 gates would be certainly beyond the scope of our presentation. We therefore selected gates representing five characteristic cellular subsets, namely gates 1, 3, 6, 12 and 26, to illustrate the possibilities, but also the potential shortcomings associated with a t-SNE representation. The corresponding box plots are shown in Figure 4.
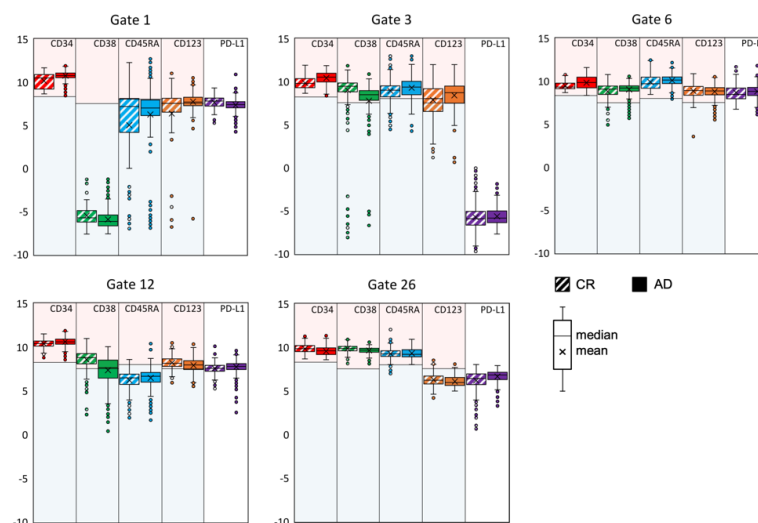
Figure 4. Box plots of the antigen expression levels (red: CD34, green: CD38, blue: CD45RA, orange: CD123, purple: PD-L1) within 5 selected gates from the 27 shown in Figure 3. The fluorescence data were scaled biexponentially in a preliminary step. The light blue and red background indicate the expression levels classified as *negative* (blue) and *positive* (red), respectively. The PD-L1 antigen was not used for the cell classification.

In general, a clear distinction of one gate from the others originates from a particularly low expression of one antigen within this gate.

We begin with gate 1, a well-separated island containing the great majority of HSC/MPP cells, as defined by the lack, or extremely low expression, of CD38. Since the differentiation between HSC and MPP is based on the CD90 marker (with CD38 negative in both cases) which was not included in our panel, we cannot distinguish these two cell types within our data set. As far as CD45RA and CD123 are concerned, their expression levels show a broad distribution spanning almost the full intensity range. Since AD cells contribute 88% to this population, this gate represents a predominantly leukemic-related gate and is compatible with the signature of leukemic stem cells. We note that the patients' ID and the group assignment were added to the expression matrix prior to the data merging, which allows us to determine the contribution of each patient group (CR and AD) to each gate. To relate these findings to the results of Kersten et al. [32], we looked at the expression level of CD45RA and CD123 on the CD34$^+$ cells within this gate and found a greater expression of these antigens on the leukemic cells compared to those from the control samples. The aforementioned investigators examined the potency of CD45RA to specifically discriminate LSC and normal HSC for a better LSC quantification and found that in comparison to other markers such as CLEC12A, CD33 and CD123, CD45RA was the most reliable antigen. From a clinical point of view, it was interesting to note that CD45RA$^+$ LSC tended to be associated with a more favorable cytogenetic/molecular marker constellation. However, it is important to recognize that the expression of CD45RA in AML is not as straightforward as in the immune system T cell subsets, and the functional implications can be quite diverse [33]. With regard to CD123, the study by Testa et al. based on the screening of CD123 expression in various hematopoietic malignancies shows that this antigen not only frequently expressed at high levels in AMLs but also on B-ALLs [34]. In an earlier report, they had explored a large set of AML patients and reported that 45% of these patients overexpress CD123 [35]. Similar to their results, Al-Mawali et al. [36] found that overall, this antigen was expressed in 37 (97%) out of 38 AML cases analyzed. The median expression of CD123 was 90% (range 21%–99%). Interestingly, the proportion of cells co-expressing CD123 on CD34$^+$/CD38$^-$ leukemic stem cells was also 37 (97%) out of the 38 AML patients with a broad range from 0.0262% to 39.7% (median 0.8164, mean 4.45)

at the time of diagnosis. These results are in line with our findings regarding the expression pattern of the CD34$^+$/CD38$^-$ in our gate 1.

Gate 3, on the other hand, has been selected as an example for a cell cluster mainly encompassing CD34$^+$ normal progenitor cells of GMP subtype, as the great majority of cells show a strong CD38 expression in the presence of CD45RA and CD123. Different from this normal signature, the few CD34$^+$ cells falling onto this gate from patients with AD are lacking or only faintly expressing CD38 while the intensity for CD45RA and CD123 tends to be stronger in comparison to their normal counterparts.

Gate 6 resides at the edge of the main island with a proportion of 93% of cells from AD samples. Since the expression levels of all antigens are above the threshold of detection, they are formally classified as GMP. Still, a specific property of gate 6 in comparison to other gates containing GMP-like cells is that the PD-L1 expression level is relatively high—well above the levels in all other gates—and the levels of CD45RA and CD38 are also above the average observed for GMP cells. Furthermore, it is remarkable that these cells have a relatively low CD34 antigen expression and that all antigens display a relatively sharp intensity distribution with relatively low standard deviations. This suggests that there is no ongoing evolution among the cells in this gate. The CD34$^+$ cells of this cluster were to some extent CD38$^+$, indicating a kind of "late" HSC on its way towards an abnormal stage of differentiation. As far as PD-L1 is concerned, our t-SNE-based data confirm the results obtained previously in a study focusing on the immunophenotype of T cells in patients with MDS and AML [37]. The mechanisms underlying T cell evasion to immune checkpoint inhibitors in acute myeloid leukemia have been recently elucidated by Gurska et al. [38].

We now take a closer look at the cells in gate 12, where two-thirds of the cells originate from AD samples. This gate represents a kind of borderline cell pool regarding the AD samples. In general, the expression level of CD45RA is very low, and the CD34 level is extraordinarily high with a relatively broad distribution of CD38 expression. While the cells in this gate from the CR patients are unequivocally classified as CMP, this is not possible for the AD samples, as they rather appear to be a mixture of CMP with HSC/MPP. This gate is, therefore, distinct from most other gates due to its internal shift of the t-SNE intensity between AD and CR samples. Accordingly, the AD cells with a lack or very low expression of CD38 reside more at the left side of this gate, whereas the cells of the CR samples preferentially group around its center. This indicates that the cells undergo an evolution from HSC/MPP when the disease is active, towards CMP during remission. The cluster contained within gate 12 is thus a nice example for the discriminative strength of t-SNE. In comparison to gate 1, the CR group shows a positive CD38 signal, while the AD group in this gate is CD38 negative, even though these values are significantly higher than in gate 1.

Gate 26 is dominated by a proportion of 68% of CR cells. It is a kind of enigmatic cluster, as this subpopulation of CD34$^+$ cells could not be allocated unequivocally according to the classification scheme as described in Figure 3F. Their characterization certainly requires an extended labelling for the lymphoid progenitor cells including antigen markers like CD10, CD7 or CD19. With regard to the leukemic cells contained within this cluster, aberrant marker constellations not related to the canonic scheme are also conceivable. Therefore, starting from our proof-of-principle marker panel, modifications including new monoclonal antibodies are necessary taking into account the steadily evolving knowledge and discovery of leukemic-related antigens and their co-expression patterns. Within this process, our efforts should be geared towards linking the phenotypical characterization to the molecular signature of the leukemic cells in the sense of a phenotype–genotype linkage. In the context of an antigen-targeted therapy, this could be helpful in defining the most relevant subset, i.e., leukemic stem cell, within the bulk mass of leukemic cells.

We proceed by drawing some general conclusions from these characteristic examples.

First, carefully selected additional markers can discriminate the cells to a deeper level. In that respect, we found a strong correlation between the expression level of the PD-L1 antigen and the percentage of predominantly leukemic cells in a particular

gate. Considering that PD-L1 is an immunoprotective antigen, one may speculate that by increasing the PD-L1 expression during the evolution from healthy towards malignant, the cells protect themselves with respect to the immune system.

On the other hand, disregarding a relevant marker can leave the cell population within some of the gates unspecified, as has been seen from the example of gate 26. Moreover, since the markers used tend to show a continuous expression on this cell ensemble, only a few distinct islands became apparent in the t-SNE plot. This means that with manual density-based gating, the areas sometimes do not have a distinct border, which is reflected by the variability of the box plots for the respective gates. By using more markers that ideally exclude each other, better separation within the t-SNE plot [39] may improve subsequent gating or also enable the use of more automated density-based gating, such as DBSCAN [40] or HDBSCAN [41].

Furthermore, in our control samples of patients (CR), the composition of the cell ensemble was similar to our previous findings in normal donors showing a predominance of the GMP followed by the CMP, HSC and the MEP [42]. Subtle differences may be explained by the fact that in our study, BM samples of patients in CR served as normal controls, as BM from normal volunteers were not available. More specifically, normal hematopoietic cells that express high levels of CD34 lacking CD38 are considered stem cells, whereas those that express low levels of CD34 and high levels of CD38 represent more differentiated progenitor cells [43]. The lack of CD38 on leukemic blast cells is also characteristic for the leukemic stem cell [4].

3.3. Quantification of the t-SNE Representation

We proceed by asking to what extent a t-SNE-based assessment can be quantified. Based on quantitative evaluations already proposed [44,45], we suggest an analysis in terms of the Pearson correlation coefficient r (A, B), a well-established measure for the similarity of two pictures labelled A and B. The two pictures are composed of N pixels each, with pixel density A_j and B_j, respectively. The Pearson coefficient is defined as

$$r(A, B) = \frac{cov(A, B)}{\sigma_A \sigma_B} \tag{1}$$

with the covariance of the two pictures given by

$$cov(A, B) = \frac{\sum_{j=1}^{N} A_j B_j - \left(\sum_{j=1}^{N} A_j\right) \cdot \left(\sum_{j=1}^{N} B_j\right)}{N} \tag{2}$$

and the standard deviation of the pixel densities of picture X (X = A, B) given by

$$\sigma_X = \sqrt{\frac{\sum_{j=1}^{N} X_j^2 - \left(\sum_{j=1}^{N} X_j\right)^2}{N}} \tag{3}$$

For r (A, B) = 1, the two pictures are identical, and they are maximally different, i.e., their sum picture has a density of zero at all sites, for r (A, B) = −1.

The comparison of the two representations of the combined data sets, Figure 2B,C gives $r_{\sum CR, \sum AD} = 0.46$. Here, $\sum AD$ and $\sum CR$ denote the sum pictures of all AD and all CR samples, respectively.

Based on this value, we evaluate a classification protocol in which the t-SNE representation of a new sample is generated by first merging it with a reference plot composed of a sufficient number of samples, which is split up again into the two reference pictures $\sum AD$ and $\sum CR$ plus the contribution from the new sample labelled as N.

When we refer to the sample N including its classification, we label it by NCR or NAD, respectively. In the next step, the Pearson coefficients of N with the two reference pictures $r_{\sum AD, N}$ and $r_{\sum CR, N}$ are computed.

We have implemented this protocol with the present data set as follows. From our data set, we have removed each sample separately and considered the remaining combined pictures as reference pictures. We now treat the individual sample N as unknown and compute $r_{\sum AD,N}$ as well as $r_{\sum CR,N}$. This implementation is represented schematically in Figure 5.

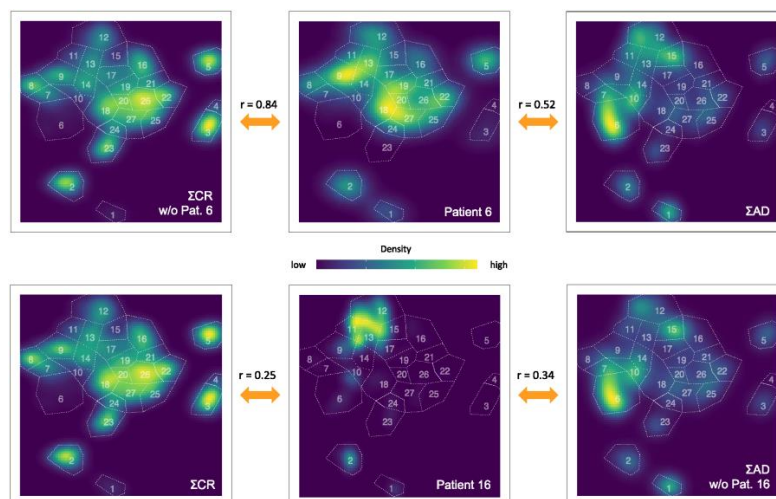**Figure 5.** Graphic representation of the calculation of the r value as an example for two patients, 6 (CR group, top center) and 16 (AD group, bottom center). For patient 6, the density distribution is compared with those containing all patients with AD (top right) and for all patients of the CR group except patient 6 (top left). Likewise, we compare for patient 16 the density with the t-SNE plots containing all patients with AD except patient 16 (bottom right) and all patients of the CR group (bottom left). The orange arrows indicate which two plots were compared with each other. The r values are given in between the compared plots. The density is normalized to the respective maximum value of the plot. The gates are shown as an overlay for all plots.

The results are the values without parentheses listed in Table 4. For all but one of the twelve control samples, we measure $r_{\sum CR,NCR} > r_{\sum AD,NCR}$, with differences up to 0.47 for sample 20. Therefore, only in the case of $N = 18$, the sample would have been classified as AD in contrast to the correct classification. We will elucidate the reasons for the wrong classification below.

Regarding the identification of an AD sample, the situation is less clear. While samples 1, 12, 13, 14 and 16 show $r_{\sum AD,NAD} > r_{\sum CR,NAD}$ and are thus classified correctly, we observe that $r_{\sum AD,NAD}$ is just slightly smaller than $r_{\sum CR,NAD}$ in samples 10 and 17, but find dramatic deviations from the classification for samples 4 and 11, with r value differences of 0.68 and 0.42, respectively.

In order to investigate how stably the classification works with respect to multiple t-SNE runs, two further runs were performed, and the classification was carried out as described previously. The AD samples were assigned to the same group in all runs as described above. In two out of three runs, all CR samples except $N = 18$ were identified as CR samples and in the third run, all were classified as CR.

The failure of allocating samples 4 and 11 asks for refined consideration. Despite their different subtype of AML, the leukemic cells show a monoblastic differentiation reflecting a more "mature" subtype not necessarily reflected by a particular CD34/CD38 subset. Since the cells of these misassigned patients represent a more mature type and the patients had only a molecular relapse, it is very likely that they could not be adequately assigned, since the leukemic cells were not contained within the CD34$^+$ cell population. For the detection of that kind of subtype, additional markers such as CD33 and CD14, for example, would still

be necessary. Rather, the antigen markers used should show expression levels quite similar to those of the CR samples. Since it is of great interest to study how such a misallocation influences the t-SNE representation and the corresponding Pearson coefficients, we remove patients 4 and 11 from the ensemble and repeat the quantitative analysis. The modified t-SNE plot in comparison to the plots of these two patients, shown in Figure 6, illustrates the dissimilarity of the density distributions. The obtained r values are given in Table 4 in parentheses. First, we notice a striking decrease of $r_{\sum AD, \sum CR}$ by 0.22. Apparently, these two samples have been responsible for a significant similarity between the two t-SNE representations, again indicating that samples 4 and 11 generate a pattern that resembles more CR samples than AD cases. Second, for all control samples, the values of $r_{\sum AD, NCR}$ improve, some of them dramatically, e.g. for patient 9, r drops from 0.24 to −0.05. Third, however, we observe some effect on the $r_{\sum AD, NAD}$ values, which change by no more than 0.19. It increases only for patients 1 and 13 but decreases for the remaining cases. This impressively shows how the lack of a relevant marker for clear characterization can lead to false similarities and thus impede the classification. It is therefore conceivable that with each additional diagnostically relevant marker the characterization becomes better, the t-SNE image becomes more differentiated and thus the classification becomes more reliable.

Table 4. The calculated r values are shown, in the second respective third column, the density of the t-SNE plot of the individual patient (first column) with AD is compared with the density of the t-SNE plot containing all patients with AD ($\sum$AD) respective of the density of the t-SNE plot containing all patients of the CR group ($\sum$CR). In columns 4 to 6, this is shown accordingly for the individual patients from the CR group (fourth column), in each case compared to the density of the t-SNE plot of the entire CR group ($\sum$CR) or entire group of patients with AD ($\sum$AD). In the seventh column, the r value for the comparison between the densities of the two t-SNE plots with all patients from the CR group ($\sum$CR) and with all patients with AD ($\sum$AD) is shown. The calculated r values from the analysis in which patients 4 and 11 were excluded from the AD group are given in parentheses in the respective column.

1	2	3	4	5	6	7
Pat.	NAD vs. $\sum$AD	NAD vs. $\sum$CR	Pat.	NCR vs. $\sum$CR	NCR vs. $\sum$AD	$\sum$CR vs. $\sum$AD
1	0.23 (0.25)	0.17	2	0.46	0.17 (0.00)	0.46 (0.24)
4	0.12	0.80	3	0.53	0.41 (0.36)	
10	−0.01 (−0.2)	0.06	5	0.77	0.44 (0.23)	
11	0.29	0.71	6	0.84	0.52 (0.31)	
12	0.14 (0.13)	0.12	7	0.50	0.43 (0.41)	
13	0.05 (0.07)	−0.06	8	0.56	0.19 (0.02)	
14	0.22 (0.19)	0.14	9	0.61	0.24 (-0.05)	
16	0.34 (0.33)	0.25	15	0.70	0.55 (0.43)	
17	0.12 (0.11)	0.15	18	0.43	0.48 (0.42)	
			19	0.28	0.18 (0.16)	
			20	0.67	0.25 (0.01)	
			21	0.69	0.43 (0.28)	

As an evaluation of this proposed identification protocol, we note that the values of $r_{\sum CR, NCR}$ are large, a fact which quantifies the high similarity of the cell population in the CR stage. They are furthermore significantly larger than $r_{\sum AD, NCR}$ and we can thus conclude that the state of CR is safely identified and clearly distinguished from the AD state. Furthermore, it can be characterized by a single number with the t-SNE-based protocol, namely by $r_{\sum CR, N}$. The identification of an AD case, however, has remained ambiguous. All values for $r_{\sum AD, NAD}$ and $r_{\sum CR, NAD}$ are close to zero, with correspondingly small differences which in some cases would even indicate a remission. This situation reflects, in our opinion, the heterogeneity of the considered AD cases. Since these samples generate widely varying t-SNE patterns, they have relatively low r values and if such a reference pattern is compared with a new sample, a t-SNE-based identification is ambiguous if not impossible. On the other hand, we have seen in the example of patients 4 and 11 how the t-SNE-based identification can be improved considering blasts of a more "mature"

type. We therefore expect that sufficiently differentiated t-SNE reference maps for AD subtypes will also allow the unique identification of an AD as well as its predominant blast population. To be on the safe side, we estimate that t-SNE subtype reference maps should be constructed from at least ten samples.
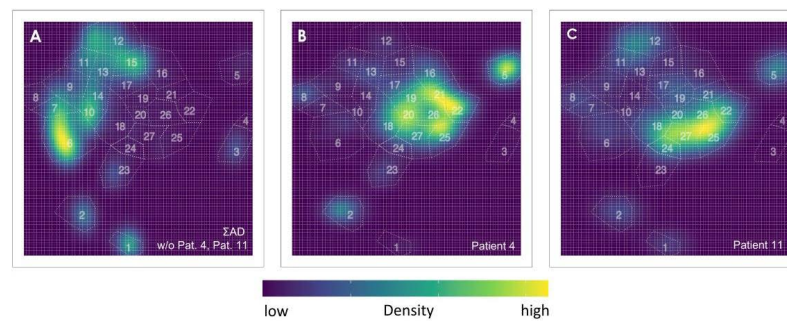

Figure 6. (**A**) Density plot of all patients with AD without patient 4 and patient 11; (**B**) density plot of patient 4, and (**C**) of patient 11. The gates are shown as an overlay in all plots.

4. Conclusions

Our study has evaluated the potential of t-SNE to represent multi-dimensional cell ensemble data from AML and MDS patients in a compact, two-dimensional form, thereby condensing the widely spread information of the scatter plots in a single picture. In order to develop this mapping into a diagnostically valuable tool, the mapping has to be capable of handling the specific challenges of such data sets, namely the large variance of the cell numbers per sample, the diverse manifestations of the diseases and their subtypes, as well as the unavoidable smearing of the map by the continuous cell evolution. Our protocol takes these initial conditions into account providing a meaningful clustering with gates containing diagnostically relevant cell populations. Additional markers may facilitate further dissection of otherwise homogeneous cell populations or hidden subtypes including rare cells not detectable in the two-dimensional scatter plots. We have also demonstrated how new samples may be diagnosed with the help of reference t-SNE patterns, based on similarities or dissimilarities, respectively. A quantified approach may comprise a statistical measure of the similarity of two pictures, like the Pearson coefficient. Our study shows that such an approach may work even for relatively poorly defined reference pictures. It is straightforward to adapt this concept to the evolution of the cell population of individual patients under therapy.

To facilitate the continued advancement of the t-SNE method, it is essential to establish a consistent and unchanging assignment of cells to a predefined t-SNE map representing "normality," as detailed earlier. This reference map should remain constant regardless of the introduction of new samples. Recent developments by Kobak, D. and Berens, P. [21] and Policar et al. [46] have introduced methods for embedding new samples into an existing t-SNE plot in single-cell transcriptomics data. The integration of corresponding tools in the mapping algorithm was, however, beyond the scope of the present work and will be the topic of future studies. Finally, it should be emphasized that these concepts are not limited to AML/MDS cells but can be applied to essentially all multidimensional diagnostic fluorescence flow cytometry data.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers16071320/s1, Figure S1: Evaluation of the impact of varying the t-SNE parameters number of iterations and perplexity; Figure S2: Scatter plots to determine the classification limits; Figure S3: Box plots with antigen expression levels of all 27 gates; Figure S4: Density-plots of all patients; Figure S5: t-SNE plots of the common datasets of the CR and AD group; Figure S6: t-SNE plots as an example for four patients; Table S1: Percentage distribution of the calls in the 27 gates; Table S2: MRD

analysis; Code S1: R script for the t-SNE analysis, calculation of the density matrices and some figures; Code S2: Python script for quantitative analysis of the t-SNE plots with the Pearson coefficient.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethics Committee of the University Hospital of the Heinrich-Heine-University (2020–1222) on 8 December 2020.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent for publication was not explicitly asked for, as the participating patients agreed on the exploitation of the scientific findings for the public in general.

Data Availability Statement: Data can be provided by the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AD	active disease
AML	acute myeloid leukemia
BM	bone marrow
CLP	common lymphoid progenitors
CMP	common myeloid progenitors
CR	complete remission
GMP	granulocyte-macrophage progenitors
GOI	gate of interest
HSCs	hematopoietic stem cells
HSPC	hematopoietic stem and progenitor cell
HSPCs	hematopoietic stem and progenitor cells
LSC	leukemic stem cell
LSCs	leukemic stem cells
MDS	myelodysplastic syndromes
MEF	marker expression function
MEP	megakaryocyte/erythroid progenitors
MPP	multipotent progenitor cells
PB	peripheral blood
PCA	principal component analysis
PD-L1	programmed death ligand 1
t-SNE	t-distributed stochastic neighbor embedding
UMAP	uniform manifold approximation and projection

References

1. Bonnet, D.; Dick, J.E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **1997**, *3*, 730–737. [CrossRef] [PubMed]
2. Döhner, H.; Wei, A.H.; Appelbaum, F.R.; Craddock, C.; DiNardo, C.D.; Dombret, H.; Ebert, B.L.; Fenaux, P.; Godley, L.A.; Hasserjian, R.P.; et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood* **2022**, *140*, 1345–1377. [CrossRef]

3. Tran, T.B.; Siddon, A.J. Molecular findings in myeloid neoplasms. *Int. J. Lab. Hematol.* **2023**, *45*, 442–448. [CrossRef] [PubMed]
4. Lapidot, T.; Sirard, C.; Vormoor, J.; Murdoch, B.; Hoang, T.; Cacerescortes, J.; Minden, M.; Paterson, B.; Caligiuri, M.A.; Dick, J.E. A cell initiating human acute myeloid-leukemia after transplantation into SCID mice. *Nature* **1994**, *367*, 645–648. [CrossRef]
5. Niu, J.L.; Peng, D.Y.; Liu, L.B. Drug Resistance Mechanisms of Acute Myeloid Leukemia Stem Cells. *Front. Oncol.* **2022**, *12*, 896426. [CrossRef] [PubMed]
6. Adan, A.; Alizada, G.; Kiraz, Y.; Baran, Y.; Nalbant, A. Flow cytometry: Basic principles and applications. *Crit. Rev. Biotechnol.* **2017**, *37*, 163–176. [CrossRef] [PubMed]
7. Liechti, T.; Roederer, M. OMIP-051-28-color flow cytometry panel to characterize B cells and myeloid cells. *Cytom. A* **2019**, *95*, 150–155. [CrossRef] [PubMed]
8. Fokken, H.; Waclawski, J.; Kattre, N.; Kloos, A.; Müller, S.; Ettinger, M.; Kacprowski, T.; Heuser, M.; Maetzig, T.; Schwarzer, A. A 19-color single-tube full spectrum flow cytometry assay for the detection of measurable residual disease in acute myeloid leukemia. *Cytom. A* **2023**, *105*, 181–195. [CrossRef] [PubMed]
9. Pyne, S.; Hu, X.L.; Wang, K.; Rossin, E.; Lin, T.I.; Maier, L.M.; Baecher-Allan, C.; McLachlan, G.J.; Tamayo, P.; Hafler, D.A.; et al. Automated high-dimensional flow cytometric data analysis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8519–8524. [CrossRef]
10. Verschoor, C.P.; Lelic, A.; Bramson, J.L.; Bowdish, D.M.E. An introduction to automated flow cytometry gating tools and their implementation. *Front. Immunol.* **2015**, *6*, 380. [CrossRef]
11. Moon, K.R.; van Dijk, D.; Wang, Z.; Gigante, S.; Burkhardt, D.B.; Chen, W.S.; Yim, K.; van den Elzen, A.; Hirn, M.J.; Coifman, R.R.; et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **2019**, *37*, 1482–1492. [CrossRef]
12. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
13. Diggins, K.E.; Ferrell, P.B.; Irish, J.M. Methods for discovery and characterization of cell subsets in high dimensional mass cytometry data. *Methods* **2015**, *82*, 55–63. [CrossRef]
14. Saeys, Y.; Van Gassen, S.; Lambrecht, B.N. Computational flow cytometry: Helping to make sense of high-dimensional immunology data. *Nat. Rev. Immunol.* **2016**, *16*, 449–462. [CrossRef]
15. Mair, F.; Hartmann, F.J.; Mrdjen, D.; Tosevski, V.; Krieg, C.; Becher, B. The end of gating? An introduction to automated analysis of high dimensional cytometry data. *Eur. J. Immunol.* **2016**, *46*, 34–43. [CrossRef]
16. Hu, Z.C.; Bhattacharya, S.; Butte, A.J. Application of Machine Learning for Cytometry Data. *Front. Immunol.* **2022**, *12*, 787574. [CrossRef]
17. Amir, E.D.; Davis, K.L.; Tadmor, M.D.; Simonds, E.F.; Levine, J.H.; Bendall, S.C.; Shenfeld, D.K.; Krishnaswamy, S.; Nolan, G.P.; Pe'er, D. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **2013**, *31*, 545–552. [CrossRef]
18. McInnes, L.; Healy, J.; and Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
19. Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.A.; Kwok, I.W.H.; Ng, L.G.; Ginhoux, F.; Newell, E.W. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **2019**, *37*, 38–44. [CrossRef]
20. Cheung, M.; Campbell, J.J.; Whitby, L.; Thomas, R.J.; Braybrook, J.; Petzing, J. Current trends in flow cytometry automated data analysis software. *Cytom. A* **2021**, *99*, 1007–1021. [CrossRef]
21. Kobak, D.; Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* **2019**, *10*, 5416. [CrossRef]
22. Wattenberg, M.; Viégas, F.; Johnson, I. How to use t-SNE effectively. *Distill* **2016**. Available online: https://distill.pub/2016/misread-tsne (accessed on 15 November 2023). [CrossRef]
23. Huang, H.Y.; Wang, Y.F.; Rudin, C.; Browne, E.P. Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization. *Commun. Biol.* **2022**, *5*, 719. [CrossRef]
24. Wang, Y.F.; Huang, H.Y.; Rudin, C.; Shaposhnik, Y. Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization. *J. Mach. Learn. Res.* **2021**, *22*, 1–73.
25. Kobak, D.; Linderman, G.C. Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nat. Biotechnol.* **2021**, *39*, 156–157. [CrossRef]
26. Finak, G.; Perez, J.M.; Weng, A.; Gottardo, R. Optimizing transformations for automated, high throughput analysis of flow cytometry data. *Bmc Bioinform.* **2010**, *11*, 546. [CrossRef]
27. Belkina, A.C.; Ciccolella, C.O.; Anno, R.; Halpert, R.; Spidlen, J.; Snyder-Cappione, J.E. Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nat. Commun.* **2019**, *10*, 5415. [CrossRef]
28. Civin, C.I.; Loken, M.R. Cell-surface antigens on human marrow-cells—Dissection of hematopoietic development using monoclonal-antibodies and multiparameter flow-cytometry. *Int. J. Cell Cloning* **1987**, *5*, 267–288. [CrossRef]
29. Terstappen, L.; Huang, S.; Safford, M.; Lansdorp, P.M.; Loken, M.R. Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34+CD38- progenitor cells. *Blood* **1991**, *77*, 1218–1227. [CrossRef] [PubMed]
30. Lansdorp, P.M.; Sutherland, H.J.; Eaves, C.J. Selective expression of CD45 isoforms on functional subpopulations of CD34+ hematopoietic-cells from human bone-marrow. *J. Exp. Med.* **1990**, *172*, 363–366. [CrossRef] [PubMed]
31. Sato, N.; Caux, C.; Kitamura, T.; Watanabe, Y.; Arai, K.; Bancherau, J.; Miyajima, A. Expression and factor-dependent modulation of the interleukin-3 receptor subunits on human hematopoietic-cells. *Blood* **1993**, *82*, 752–761. [CrossRef]

32. Kersten, B.; Valkering, M.; Wouters, R.; van Amerongen, R.; Hanekamp, D.; Kwidama, Z.; Valk, P.; Ossenkoppele, G.; Zeijlemaker, W.; Kaspers, G.; et al. CD45RA, a specific marker for leukaemia stem cell sub-populations in acute myeloid leukaemia. *Br. J. Haematol.* **2016**, *173*, 219–235. [CrossRef] [PubMed]

33. Brod, S.A.; Rudd, C.E.; Purvee, M.; Hafler, D.A. Lymphokine regulation of CD45R expression on human t-cell clones. *J. Exp. Med.* **1989**, *170*, 2147–2152. [CrossRef]

34. Testa, U.; Pelosi, E.; Castelli, G. CD123 as a Therapeutic Target in the Treatment of Hematological Malignancies. *Cancers* **2019**, *11*, 1358. [CrossRef]

35. Testa, U.; Riccioni, R.; Militi, S.; Coccia, E.; Stellacci, E.; Samoggia, P.; Latagliata, R.; Mariani, G.; Rossini, A.; Battistini, A.; et al. Elevated expression of IL-3Rα in acute myelogenous leukemia is associated with enhanced blast proliferation, increased cellularity, and poor prognosis. *Blood* **2002**, *100*, 2980–2988. [CrossRef] [PubMed]

36. Al-Mawali, A.; Pinto, A.D.; Al-Zadijali, S. CD34+CD38–CD123+ Cells are present in virtually all Acute Myeloid Leukaemia Blasts: A promising single unique phenotype for minimal residual disease detection. *Acta Haematol.* **2017**, *138*, 75–181. [CrossRef]

37. Moskorz, W.; Cosmovici, C.; Jäger, P.S.; Cadeddu, R.P.; Timm, J.; Haas, R. Myelodysplastic syndrome patients display alterations in their immune status reflected by increased PD-L1-expressing stem cells and highly dynamic exhausted T-cell frequencies. *Br. J. Haematol.* **2021**, *193*, 941–945. [CrossRef]

38. Gurska, L.; Gritsman, K. Unveiling T cell evasion mechanisms to immune checkpoint inhibitors in acute myeloid leukemia. *CDR* **2023**, *6*, 674–687. [CrossRef] [PubMed]

39. Eshghi, S.T.; Au-Yeung, A.; Takahashi, C.; Bolen, C.R.; Nyachienga, M.N.; Lear, S.P.; Green, C.; Mathews, W.R.; O'Gorman, W.E. Quantitative Comparison of Conventional and t-SNE-guided Gating Analyses. *Front. Immunol.* **2019**, *10*, 1194. [CrossRef]

40. Ester, M.; Krigel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2 August 1996; pp. 226–231.

41. Campello, R.J.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 160–172.

42. Bruns, I.; Cadeddu, R.P.; Brueckmann, I.; Fröbel, J.; Geyh, S.; Büst, S.; Fischer, J.C.; Roels, F.; Wilk, C.M.; Schildberg, F.A.; et al. Multiple myeloma-related deregulation of bone marrow-derived CD34+ hematopoietic stem and progenitor cells. *Blood* **2012**, *120*, 2620–2630. [CrossRef]

43. Terstappen, L.; Huang, S. Analysis of bone-marrow stem-cell. *Blood Cells* **1994**, *20*, 45–63.

44. Ho, T.C.; LaMere, M.; Stevens, B.M.; Ashton, J.M.; Myers, J.R.; O'Dwyer, K.M.; Liesveld, J.L.; Mendler, J.H.; Guzman, M.; Morrissette, J.D.; et al. Evolution of acute myelogenous leukemia stem cell properties after treatment and progression. *Blood* **2016**, *128*, 1671–1678. [CrossRef]

45. Roca, C.P.; Burton, O.T.; Neumann, J.; Tareen, S.; Whyte, C.E.; Gergelits, V.; Veiga, R.V.; Humblet-Baron, S.; Liston, A. A cross entropy test allows quantitative statistical comparison of t-SNE and UMAP representations. *Cell Rep. Methods* **2023**, *3*, 100390. [CrossRef] [PubMed]

46. Policar, P.G.; Strazar, M.; Zupan, B. Embedding to reference t-SNE space addresses batch effects in single-cell classification. *Mach. Learn.* **2023**, *112*, 721–740. [CrossRef]

Characterization of CD34+ Cells from Patients with Acute Myeloid Leukemia (AML) and Myelodysplastic Syndromes (MDS) using a t-distributed stochastic neighbor embedding (t-SNE) protocol

Cathrin Nollmann, Wiebke Moskorz, Christian Wimmenauer, Paul S. Jäger, Ron P. Cadeddu, Jörg Timm, Thomas Heinzel and Rainer Haas
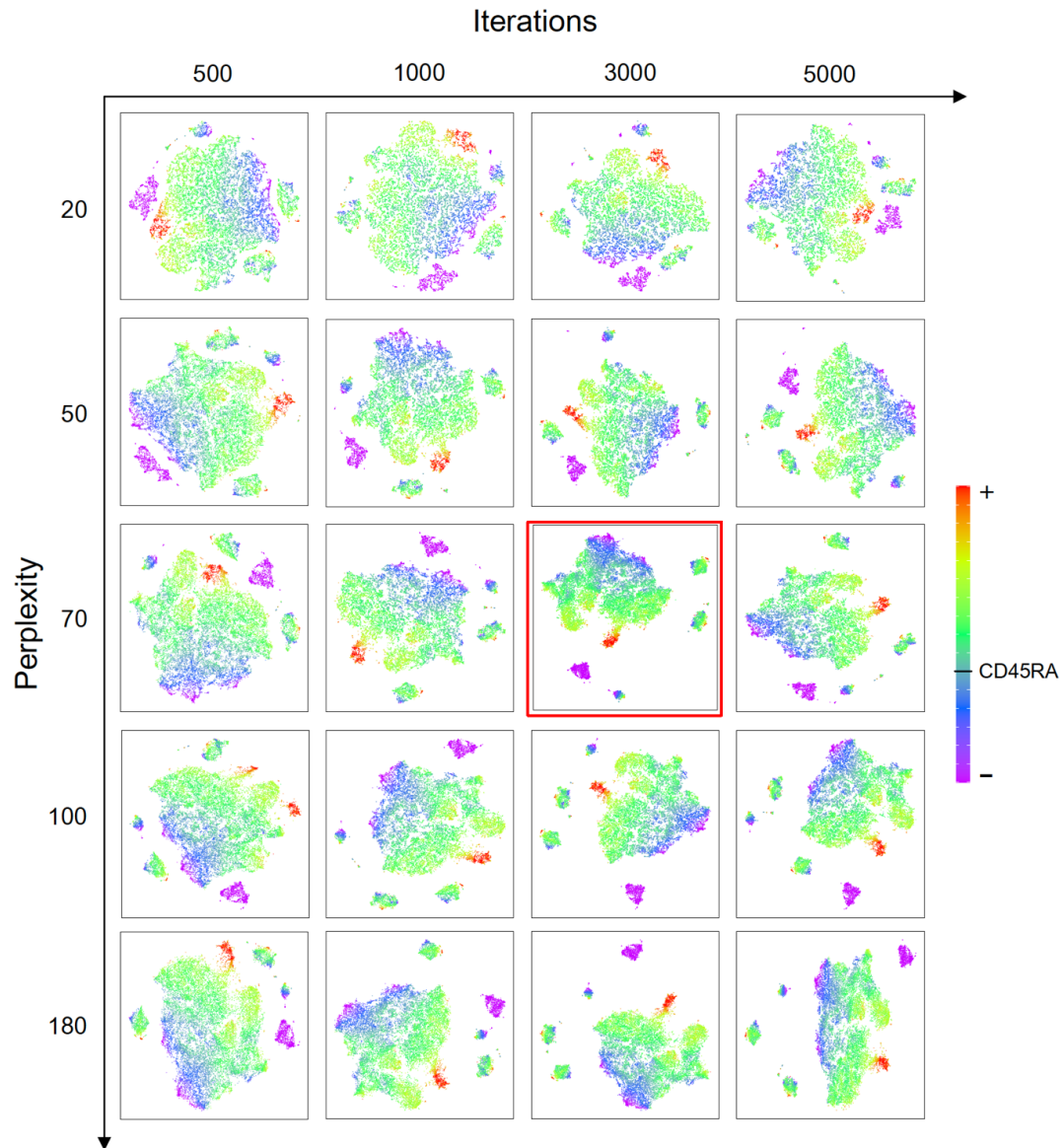

Figure S1. Evaluation of the impact of varring the t-SNE parameters number of iterations and perplexity on the t-SNE embedding. The whole dataset was analyzed. The parameter combination used for the analysis presented in the manuskript is framed in red. As an example, the expression of CD45RA in color code (color scale in arbitrary units).
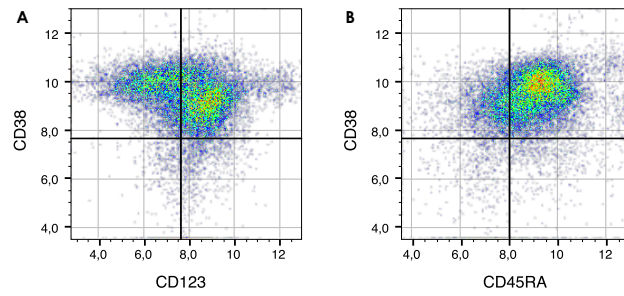
Figure S2. Scatter plots to determine the limits for the classification into *positive* and *negative* for the markers CD38 and CD123 (A) as well as for CD45RA (B). The fluorescence data was scaled biexponentially in a preliminary step.
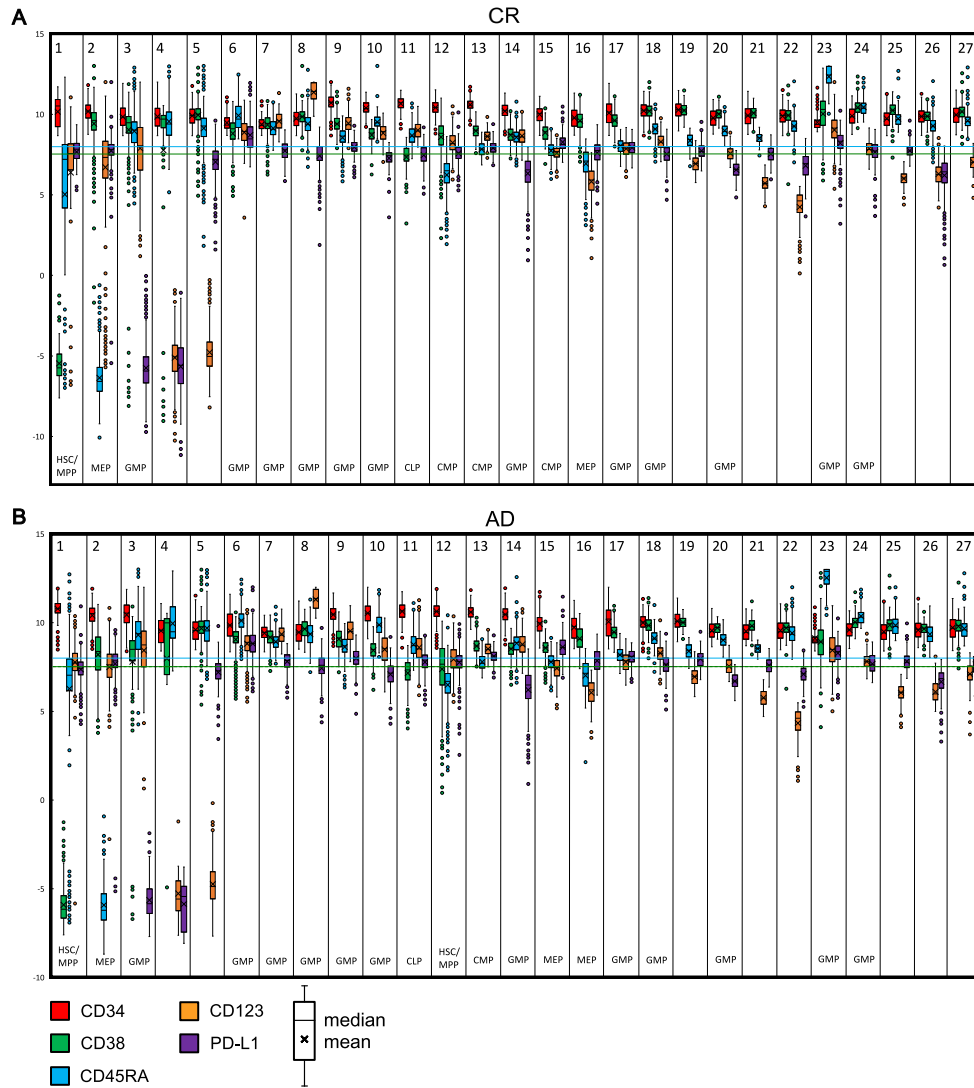
Figure S3. Box plots of the antigen expression levels of all 27 gates for (A) the patients in complete remission (CR) and (B) the patients with active disease (AD). The colored horizontal lines indicate the limits for the classification into *positive* and *negative* (blue line: CD45RA. green line: CD38 and CD123). The PD-L1 antigen was not used for the cell classification. The cell types (Table 2) assigned based on the fluorescence values are indicated at the bottom of the column: hematopoietic stem cells (HSC), multipotent progenitor cells (MPP), common lymphoid progenitors (CLP), common myeloid progenitors (CMP), megakaryocyte/ erythroid progenitors (MEP), granulocyte-macrophage progenitors (GMP).
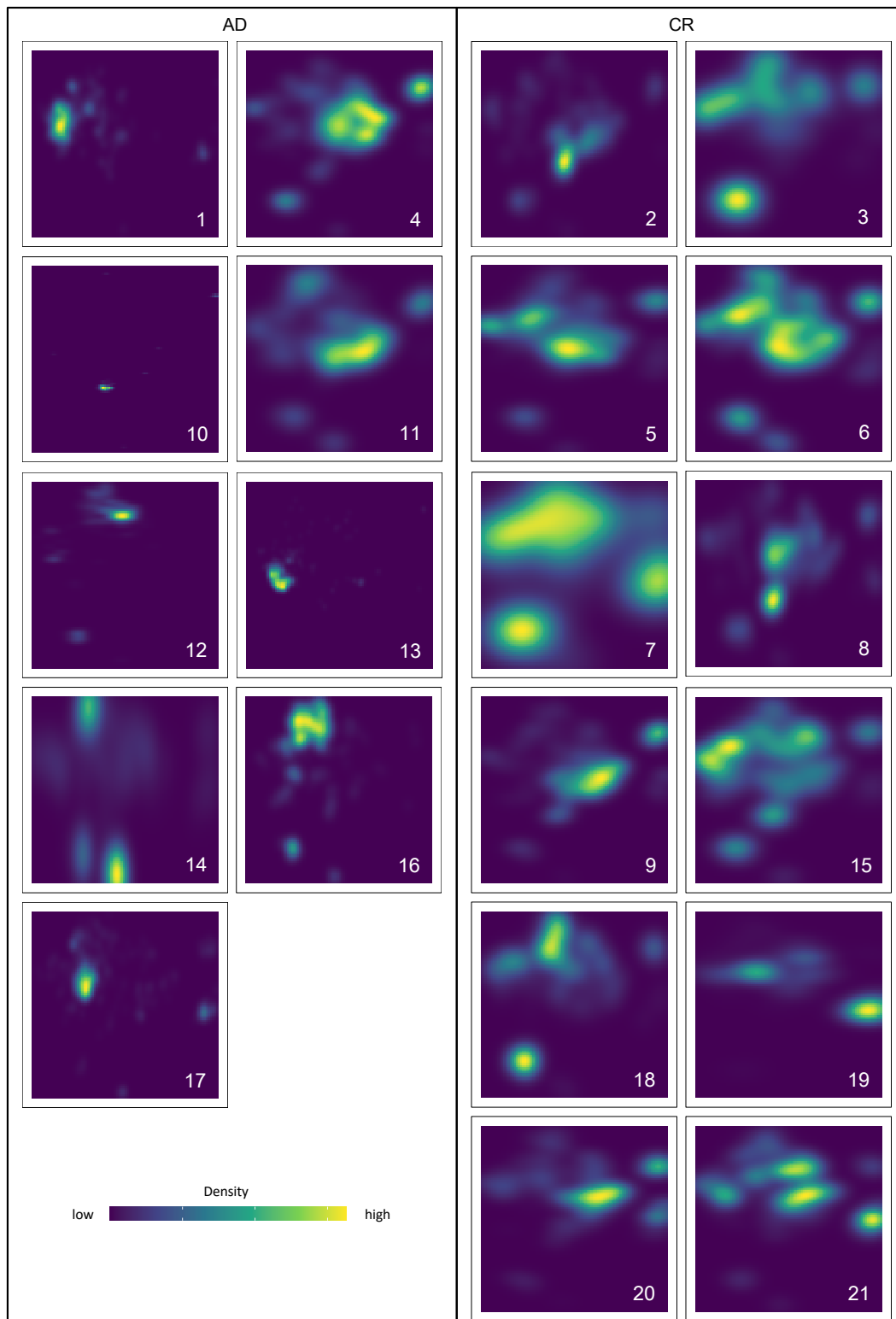
Figure S4. Density-plots of all patients, sorted according to active disease (AD, left) and complete remission (CR, right).
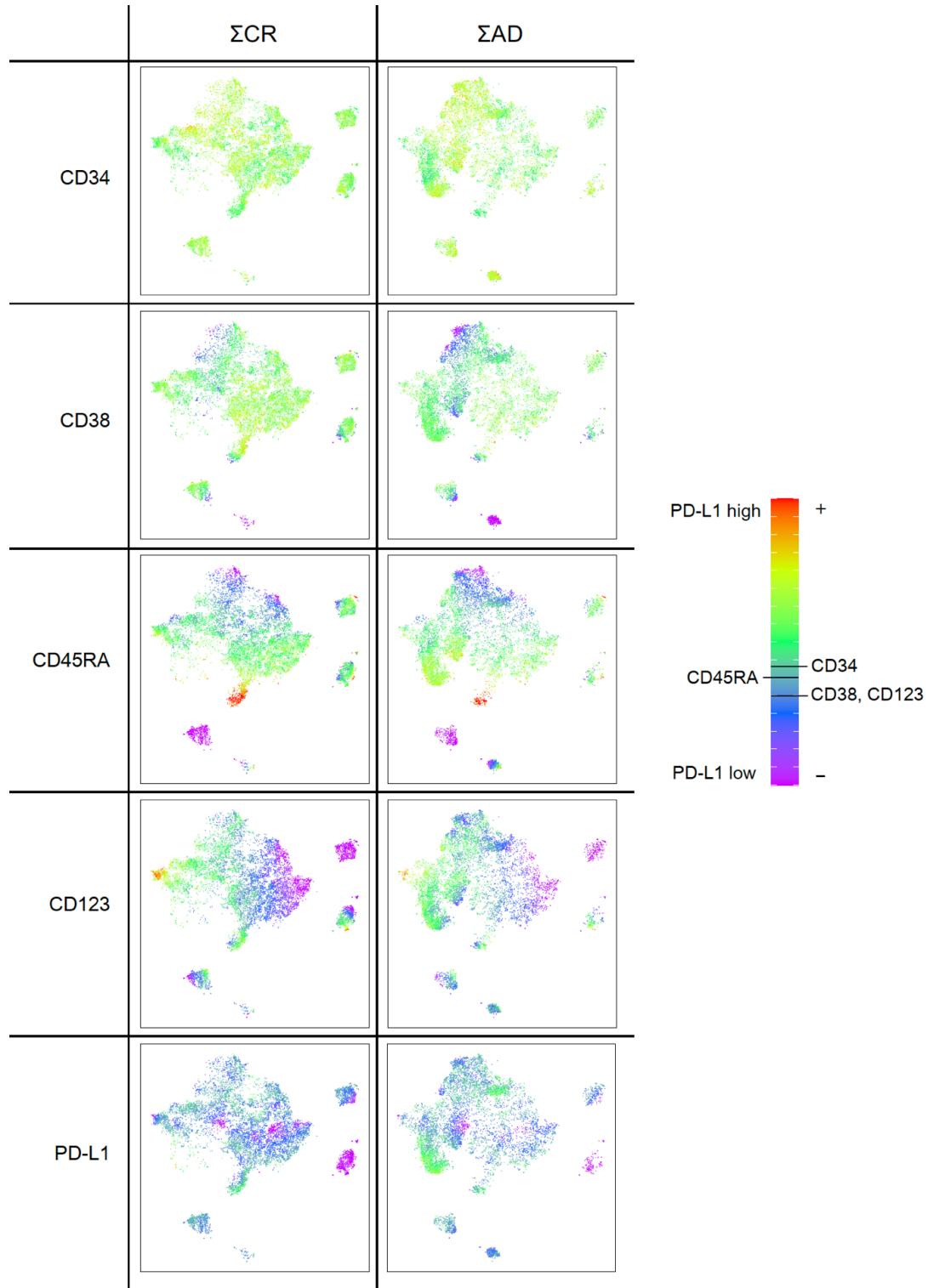
Figure S5. t-SNE represetations of the common datasets of patients in complete remission (CR) and of patients with active disease (AD)**.** The prevalence of the various markers entering the t-SNE algorithm are reproduced, for CD34, CD38, CD45RA, CD123 and PD-L1. The color scale represents the expression level in arbitrary units. The classification in terms of *positive* (+) and *negative* (-) expression is indicated by the black horizontal lines in the color bar.

4
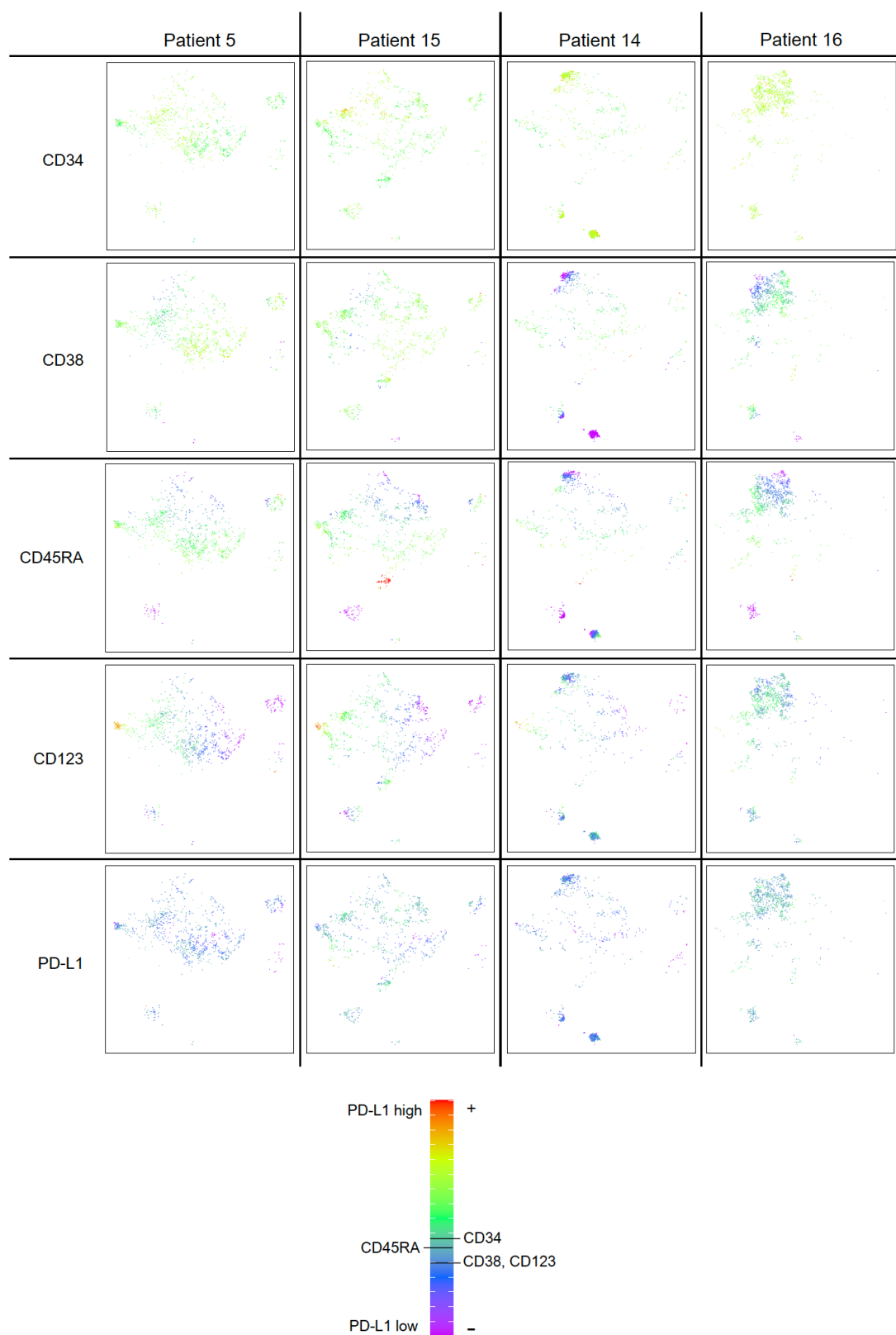
Figure S6. t-SNE represetations as an example for four patients, two each in complete remission (Patient 5, Patient 15) and with active disease (Patient 14, Patient 16). The prevalence of the various markers entering the t-SNE algorithm are reproduced, for CD34, CD38, CD45RA, CD123 and PD-L1. The color scale represents the expression level. The classification in terms of *positive* (+) and *negative* (-) expression is indicated by the black horizontal lines in the color bar.

Table S1. Percentage distribution of the cells in the 27 gates, defined in the t-SNE plot of the CR samples, for the three datasets (all patients, only AD, only CR).

Gate	All / [%]	AD / [%]	CR / [%]
1	2.50	0.65	4.79
2	5.17	5.93	4.23
3	4.64	6.88	1.87
4	0.49	0.79	0.12
5	4.86	6.35	3.01
6	9.33	1.42	19.13
7	3.37	2.87	4.00
8	2.48	3.59	1.12
9	4.10	4.98	3.00
10	2.70	0.84	5.01
11	2.49	1.22	4.06
12	6.85	4.69	9.52
13	4.10	3.60	4.73
14	3.75	3.84	3.63
15	3.93	1.15	7.36
16	3.97	5.48	2.10
17	2.71	3.43	1.82
18	5.31	7.14	3.04
19	2.59	3.51	1.45
20	2.02	2.65	1.23
21	1.57	1.83	1.26
22	3.38	4.39	2.13
23	3.57	4.71	2.16
24	1.72	2.30	1.01
25	2.91	3.20	2.55
26	5.27	7.53	2.49
27	4.22	5.05	3.18

Table S2. MRD analysis**.** All 12 patients in haematological CR were MRD negative. In two patients (7 and 9), leukaemia specific markers were monitored using next-generation sequencing (NGS) or fragment analysis, both yielding negative results. As our molecular biological methods do not detect any leukemia related mutation, MRD in the other patients is defined by a low WT1 level following an initially high expression or 100% donor chimerism.

Pat. ID	MRD markers
2	chimerism, ASXL1, WT1
3	chimerism, CEBPA, WT1
5	chimerism, WT1
6	chimerism, WT1
7	chimerism, NPM1
8	chimerism, WT1, JAK2, SF3B1
9	chimerism, CBFB-MYH11
15	chimerism, WT1, RUNX1, FISH
18	chimerism, WT1, FLT3-ITD
19	chimerism, ASXL1, WT1
20	chimerism, WT1, ASXL1, RUNX1
21	chimerism, WT1, ASXL1, RUNX1, TET2 and EZH2

Code S1. R script for the t-SNE analysis, calculation of the density matrices and some figures.

```
#LIBRARIES#################################################
library(devtools)
library(flowCore)
library(flowClean)
library(flowClust)
library(Rtsne)
library(ggplot2)
install.packages("scales")
install.packages("ggthemes")
library(viridis)
library(png)
library(reshape2)
library(MASS)
library(writexl)
library(dplyr)
###########################################################

#Transformations
BiTrans <- biexponentialTransform(b = 1, d = 1)
LinTrans <- linearTransform(a = 1/100)

#Create Dataframe
df_t <- data.frame(FSCA=numeric(0), FSCH=numeric(0), FSCW=numeric(0),
SSCA=numeric(0), SSCH=numeric(0), SSCW=numeric(0), CD34=numeric(0),
CD38=numeric(0), Viability=numeric(0), CD123=numeric(0), CD45=numeric(0),
CD45RA=numeric(0),  PDL1=numeric(0), time=numeric(0), PatientID =
numeric(0), Control = numeric(0))

#Load files and transform
patients = c(1:21)
for (patient in patients) {
  setwd(toString(patient))
  fcm <- read.FCS(paste0("00",patient,".fcs" ))
  fcm.linear = transform(fcm,`FSC-A`=LinTrans(`FSC-A`),`FSC-
H`=LinTrans(`FSC-H`),`FSC-W`=LinTrans(`FSC-W`),`SSC-A`=LinTrans(`SSC-
A`),`SSC-H`=LinTrans(`SSC-H`),`SSC-W`=LinTrans(`SSC-W`),
                        `FJComp-APC-A`=BiTrans(`FJComp-APC-A`),
                        `FJComp-APC-Cy7-A`=BiTrans(`FJComp-APC-Cy7-A`),
```

```r
                            `FJComp-AmCyan-A`=BiTrans(`FJComp-AmCyan-A`),
                            `FJComp-PE-A`=BiTrans(`FJComp-PE-A`),
                            `FJComp-PE-Cy5-A`=BiTrans(`FJComp-PE-Cy5-A`),
                            `FJComp-PE-Cy7-A`=BiTrans(`FJComp-PE-Cy7-A`),
                            `FJComp-PerCP-Cy5-5-A`=BiTrans(`FJComp-PerCP-Cy5-
5-A`) )
  df <- as.data.frame(fcm.linear@exprs)

  #Add parameter Patient
  df['Patient'] = patient

  #Add parameter control yes or no
  if(patient == 2 || patient == 3 || patient == 5 || patient == 6 ||
patient == 7 || patient == 8 || patient == 9 || patient == 15 || patient ==
18 || patient == 19 || patient == 20 || patient == 21){
    df['Control'] = 1 #Control
  } else {df['Control'] = 0} #Active Disease

  df_t <- rbind(df_t, df)
  setwd("..")
}


#Change column names
colnames(df_t) <- c("FSCA", "FSCH", "FSCW", "SSCA", "SSCH", "SSCW", "CD34",
"CD38", "Viability","CD123", "CD45", "CD45RA", "PDL1", "Time", "Patient",
"Control")

#Choose markers for t-SNE run
datamat_t <- as.matrix(df_t[c("CD34","CD38", "CD123", "CD45RA", "PDL1")])

#Run t-SNE for different parameters

iterations = c(500,1000,3000,5000)
perplexities = c(20,50,70,100,180)
markers = c("CD34", "CD38", "CD123", "CD45RA", "PDL1")

for (iteration in iterations){
  for(perp in perplexities){
    tsne <- Rtsne(datamat_t, dims = 2, perplexity=perp, verbose=TRUE,
max_iter = iteration, check_duplicates = FALSE)
    df_t[c('tSNE1', 'tSNE2')] <- as.data.frame(tsne$Y)
    perplexity <- perp
    dataname <- paste0("Data_tSNE_it",iteration,"_p",perp,".Rdata")
    save(data, file = dataname)
      for (marker in markers){
        ggplot(data=subset(df_t, ), aes_string(x="tSNE1", y="tSNE2",colour
= `marker`)) +geom_point(size = 0.1) +
scale_colour_gradientn(colours=rev(rainbow(5)), limits = c(5, 13), oob =
scales::squish) + theme_linedraw() + theme(panel.grid = element_blank(),
axis.text = element_blank(), axis.title = element_blank(), axis.ticks =
element_blank()) + guides(color = FALSE)
        ggsave(paste0(marker,"_p",perp, ".pdf"), plot = last_plot(),
              scale = 1, width = 20, height = 20, dpi = 600,
              units = "cm")
      }
  }
}
```

```r
#Create scatter plots
#Note that you have to adjust the limits according to your transformation

markers = c("CD34", "CD38", "CD123", "CD45RA", "PDL1")

#Common scatter plot all Patients, fluorescent markers color-coded
for (marker in markers){
  ggplot(data=subset(df_t,), aes_string(x="tSNE1", y="tSNE2",colour =
`marker`)) +geom_point(size = 0.1)   +
scale_colour_gradientn(colours=rev(rainbow(5)), limits = c(5, 13), oob =
scales::squish) + theme_linedraw() + theme(panel.grid = element_blank())
  ggsave(paste0(marker,"_p",perplexity, ".pdf"), plot = last_plot(),
         scale = 1, width = 22, height = 20, dpi = 600,
         units = "cm")
}

#Common scatter plot all patients with active disease, fluorescent markers
color-coded
for (marker in markers){
  ggplot(data=subset(df_t, Control==0), aes_string(x="tSNE1",
y="tSNE2",colour = `marker`)) +geom_point(size = 0.1) + theme_classic() +
scale_colour_gradientn(colours=rev(rainbow(5)), limits = c(5, 13), oob =
scales::squish) + theme_linedraw() + theme(panel.grid = element_blank())
  ggsave(paste0(marker,"_p",perplexity, "_AD", ".pdf"), plot = last_plot(),
         scale = 1, width = 22, height = 20, dpi = 600,
         units = "cm")
}

#Common scatter plot all control patients, fluorescent markers color-coded
for (marker in markers){
  ggplot(data=subset(df_t, Control==1), aes_string(x="tSNE1",
y="tSNE2",colour = `marker`)) +geom_point(size = 0.1) + theme_classic() +
scale_colour_gradientn(colours=rev(rainbow(5)), limits = c(5, 13), oob =
scales::squish) + theme_linedraw() + theme(panel.grid = element_blank())
  ggsave(paste0(marker,"_p",perplexity,"_CR", ".pdf"), plot = last_plot(),
         scale = 1, width = 22, height = 20, dpi = 600,
         units = "cm")
}

#Scatter plot for every patient, fluorescent markers color-coded
for (marker in markers){
  for (patient in patients){
    ggplot(data=subset(df_t,Patient == patient), aes_string(x="tSNE1",
y="tSNE2",colour = `marker`)) +geom_point(size = 0.1)   +
scale_colour_gradientn(colours=rev(rainbow(5)), limits = c(5, 13), oob =
scales::squish) + theme_linedraw() + theme(panel.grid = element_blank())
    ggsave(paste0("Pat",patient,"_", marker, "_p",perplexity,".pdf"), plot
= last_plot(),
         scale = 1, width = 22, height = 20, dpi = 600,
         units = "cm")
  }
}

#Create density plots with overlay of the scatter plot

#Common density plot all patients
ggplot(data=subset(df_t), aes(x=tSNE1, y=tSNE2)) +
scale_x_continuous(limits =range(df_t$tSNE1)) + scale_y_continuous(limits =
range(df_t$tSNE2) )   +geom_point(size = 0.1) +
geom_density2d_filled(contour_var = "ndensity", alpha = 0.7 ) +
theme_linedraw() + theme(panel.grid = element_blank())
ggsave(paste0("Densityplot_Scatter_p",perplexity,".pdf"), plot =
last_plot(),
       scale = 1, width = 22, height = 20, dpi = 600,
```

```r
        units = "cm")

#Common density plot with overlay of the scatter plot patients with active
disease
ggplot(data=subset(df_t, Control == 0), aes(x=tSNE1, y=tSNE2)) +
scale_x_continuous(limits =range(df_t$tSNE1)) + scale_y_continuous(limits =
range(df_t$tSNE2) )  +geom_point(size = 0.1) +
geom_density2d_filled(contour_var = "ndensity",alpha = 0.7) +
theme_linedraw() + theme(panel.grid = element_blank())
ggsave(paste0("Densityplot_Scatter_AD_p",perplexity,".pdf"), plot =
last_plot(),
        scale = 1, width = 22, height = 20, dpi = 600,
        units = "cm")

#Common density plot with overlay of the Scatter plot patients control
ggplot(data=subset(df_t, Control == 1), aes(x=tSNE1, y=tSNE2)) +
scale_x_continuous(limits =range(df_t$tSNE1)) + scale_y_continuous(limits =
range(df_t$tSNE2) )  +geom_point(size = 0.1) +
geom_density2d_filled(contour_var = "ndensity",alpha = 0.7) +
theme_linedraw() + theme(panel.grid = element_blank())
ggsave(paste0("Densityplot_Scatter_CR_p",perplexity,".pdf"), plot =
last_plot(),
        scale = 1, width = 25, height = 22, dpi = 600,
        units = "cm")


#Calculate and plot density matrix
#For every single patient

for (patient in patients) {

  data_sub = subset(df_t, Patient == patient )
  density_matrix_list = kde2d(data_sub$tSNE1, data_sub$tSNE2,lims =
c(range(df_t$tSNE1), range(df_t$tSNE2)), n=90)

  rownames(density_matrix_list$z) = density_matrix_list$x
  colnames(density_matrix_list$z) = density_matrix_list$y

  density_matrix= melt(density_matrix_list$z,
id.var=rownames(density_matrix_list))
  names(density_matrix) = c("tSNE1","tSNE2","z")

  write_xlsx(density_matrix,
paste0('Density_matrix_patient',patient,'.xlsx'))

  ggplot(density_matrix, aes(tSNE1, tSNE2, z=z, fill=z)) + geom_tile() +
scale_fill_viridis_c() + theme_linedraw() + theme(panel.grid =
element_blank())
  ggsave(paste0( "Densitymatrix_patient_",patient, ".png"), plot =
last_plot(),
          scale = 1, width = 25, height = 20, dpi = 600,
          units = "cm")

}

#Common density matrix for patient groups
patients_AD = c(1,4,10,11,12,13,14,16,17)
patients_control = c(2,3,5,6,7,8,9,15,18,19,20,21)

#Common density matrix patients with active disease

data_sub = subset(df_t, Control == 0)

density_matrix_list = kde2d(data_sub$tSNE1, data_sub$tSNE2,lims =
c(range(df_t$tSNE1), range(df_t$tSNE2)), n=90)
```

```r
rownames(density_matrix_list$z) = density_matrix_list$x
colnames(density_matrix_list$z) = density_matrix_list$y


density_matrix= melt(density_matrix_list$z,
id.var=rownames(density_matrix_list))
names(density_matrix) = c("tSNE1","tSNE2","z")

write_xlsx(density_matrix, paste0('Density_matrix_AD.xlsx'))

ggplot(density_matrix, aes(tSNE1, tSNE2, z=z, fill=z)) + geom_tile() +
scale_fill_viridis_c() + theme_linedraw() + theme(panel.grid =
element_blank())
ggsave(paste0( "Densitymatrix_AD.pdf"), plot = last_plot(),
       scale = 1, width = 22, height = 20, dpi = 600,
       units = "cm")


#Common density matrix patients with active disease without patient k
for(k in patients_AD){

  data_sub = subset(df_t, Control == 0 & Patient != k )

  density_matrix_list = kde2d(data_sub$tSNE1, data_sub$tSNE2,lims =
c(range(df_t$tSNE1), range(df_t$tSNE2)), n=90)

  rownames(density_matrix_list$z) = density_matrix_list$x
  colnames(density_matrix_list$z) = density_matrix_list$y


  density_matrix= melt(density_matrix_list$z,
id.var=rownames(density_matrix_list))
  names(density_matrix) = c("tSNE1","tSNE2","z")

  write_xlsx(density_matrix, paste0('Density_matrix_AD_woPat',k,'.xlsx'))

  ggplot(density_matrix, aes(tSNE1, tSNE2, z=z, fill=z)) + geom_tile() +
scale_fill_viridis_c() + theme_linedraw() + theme(panel.grid =
element_blank())
  ggsave(paste0( "Densitymatrix_AD_woPat",k,".pdf"), plot = last_plot(),
         scale = 1, width = 22, height = 20, dpi = 600,
         units = "cm")
}

#Common density matrix patients with active disease without patient 4 and
11

  data_sub = subset(df_t, Control == 0 & Patient != 4 & Patient != 11 )

  density_matrix_list = kde2d(data_sub$tSNE1, data_sub$tSNE2,lims =
c(range(df_t$tSNE1), range(df_t$tSNE2)), n=90)

  rownames(density_matrix_list$z) = density_matrix_list$x
  colnames(density_matrix_list$z) = density_matrix_list$y


  density_matrix= melt(density_matrix_list$z,
id.var=rownames(density_matrix_list))
  names(density_matrix) = c("tSNE1","tSNE2","z")

  write_xlsx(density_matrix,
paste0('Density_matrix_AD_woPat4_woPat11.xlsx'))

  ggplot(density_matrix, aes(tSNE1, tSNE2, z=z, fill=z)) + geom_tile() +
```

```r
  scale_fill_viridis_c() + theme_linedraw() + theme(panel.grid =
element_blank())
  ggsave(paste0( "Densitymatrix_AD_woPat4_woPat11.pdf"), plot =
last_plot(),
          scale = 1, width = 22, height = 20, dpi = 600,
          units = "cm")

#Common density matrix patients with active disease without patient 4, 11
and patient k

  data_sub = subset(df_t, Control == 0 & Patient != 4 & Patient != 11 &
Patient != k  )

  density_matrix_list = kde2d(data_sub$tSNE1, data_sub$tSNE2,lims =
c(range(df_t$tSNE1), range(df_t$tSNE2)), n=90)

  rownames(density_matrix_list$z) = density_matrix_list$x
  colnames(density_matrix_list$z) = density_matrix_list$y


  density_matrix= melt(density_matrix_list$z,
id.var=rownames(density_matrix_list))
  names(density_matrix) = c("tSNE1","tSNE2","z")

  write_xlsx(density_matrix,
paste0('Density_matrix_AD_woPat4_woPat11_woPat',k,'.xlsx'))

  ggplot(density_matrix, aes(tSNE1, tSNE2, z=z, fill=z)) + geom_tile() +
scale_fill_viridis_c() + theme_linedraw() + theme(panel.grid =
element_blank())
  ggsave(paste0( "Densitymatrix_AD_woPat4_woPat11_woPat",k,".pdf"), plot =
last_plot(),
          scale = 1, width = 22, height = 20, dpi = 600,
          units = "cm")


#Common density matrix patients control

data_sub = subset(df_t, Control == 1)

density_matrix_list = kde2d(data_sub$tSNE1, data_sub$tSNE2,lims =
c(range(df_t$tSNE1), range(df_t$tSNE2)), n=90)

rownames(density_matrix_list$z) = density_matrix_list$x
colnames(density_matrix_list$z) = density_matrix_list$y


density_matrix= melt(density_matrix_list$z,
id.var=rownames(density_matrix_list))
names(density_matrix) = c("tSNE1","tSNE2","z")

write_xlsx(density_matrix, paste0('Density_matrix_CR.xlsx'))

ggplot(density_matrix, aes(tSNE1, tSNE2, z=z, fill=z)) + geom_tile() +
scale_fill_viridis_c()  + theme_linedraw() + theme(panel.grid =
element_blank())
ggsave(paste0( "Densitymatrix_CR.pdf"), plot = last_plot(),
      scale = 1, width = 22, height = 20, dpi = 600,
      units = "cm")


#Common density matrix patients control without patient k

for(k in patients_control){
```

```r
  data_sub = subset(df_t, Control == 1 & Patient != k )

  density_matrix_list = kde2d(data_sub$tSNE1, data_sub$tSNE2,lims =
c(range(df_t$tSNE1), range(df_t$tSNE2)), n=90)

  rownames(density_matrix_list$z) = density_matrix_list$x
  colnames(density_matrix_list$z) = density_matrix_list$y


  density_matrix= melt(density_matrix_list$z,
id.var=rownames(density_matrix_list))
  names(density_matrix) = c("tSNE1","tSNE2","z")

  write_xlsx(density_matrix, paste0('Density_matrix_CR_woPat',k,'.xlsx'))

  ggplot(density_matrix, aes(tSNE1, tSNE2, z=z, fill=z)) + geom_tile() +
scale_fill_viridis_c()  + theme_linedraw() + theme(panel.grid =
element_blank())
  ggsave(paste0( "Densitymatrix_CR_woPat",k,".pdf"), plot = last_plot(),
        scale = 1, width = 22, height = 20, dpi = 600,
        units = "cm")
}


#Convert dataframe in CSV file

#For all patients

datasub_csv <- subset(df_t)
write.csv(datasub_csv, paste0("csv_dataframe_all.csv"), row.names=FALSE)


#For every single patient

for(patient in patients){
  datasub_csv <- subset(df_t, Patient==patient)
  write.csv(datasub_csv, paste0("csv_dataframe_",patient,".csv"),
row.names=FALSE)
}

#All patients with active disease

datasub_csv <- subset(df_t, Control ==0)
write.csv(datasub_csv, paste0("csv_dataframe_AD.csv"), row.names=FALSE)

#All patients control

datasub_csv <- subset(df_t, Control ==1)
write.csv(datasub_csv, paste0("csv_dataframe_control.csv"),
row.names=FALSE)


#Read FSC File of a single FlowJo Gate and convert it into a CSV File

for(k in 1:27) {
  fcm_temp <- read.FCS(paste0("export_csv_dataframe_all_",k,".fcs" ),
truncate_max_range = TRUE)
  data_csv_temp <- as.data.frame(fcm_temp@exprs)
  colnames(data_csv_temp) <- c("FSCA",
"Viability","CD123","CD45","CD45RA","PDL1","FSCH","Time","Patient","tSNE1",
"tSNE2","FSCW","SSCA","SSCH","SSCW","CD34","CD38")
  data_csv_temp <- data_csv_temp[c(paste0("FSCA"), paste0("FSCH"),
paste0("FSCW"), paste0("SSCA"), paste0("SSCH"), paste0("SSCW"),"CD45",
"CD34", "CD38", "CD45RA", "CD123", "PDL1", "Time", "Patient", "tSNE1",
"tSNE2")]
```

```
  write.csv(data_csv_temp, paste0("csv_dataframe_all_gate_",k,".csv"),
row.names=FALSE)
}
```

Code S2. Python script for quantitative analysis of the t-SNE plots with the Pearson coefficient

```python
import numpy as np
import pandas as pd


#Calculate Pearson  Coefficient

def pearson_coeff(density1, density2):
    df = pd.concat([density1[density1.columns[-1]],
                    density2[density2.columns[-1]]], axis=1)
    df.columns = ['d1', 'd2']
    pearson = df.cov().iloc[1, 0] / (df.iloc[:, 0].std() * df.iloc[:,
1].std())
    return(pearson)


#Compare common dataset AD vs common dataset control

datasets = []

datasets.append(pd.read_excel("Density_matrix_AD.xlsx"))
datasets.append(pd.read_excel("Density_matrix_CR.xlsx"))

R = []

for df1 in datasets:
    tempR = []

    for df2 in datasets:
        tempR.append(pearson_coeff(df1, df2))
    R.append(tempR)

R = np.array(R)


#Compare common dataset AD / common dataset Control with single patients
NAD / NControl

dataset_AD_woN = []
dataset_control_woN = []
dataset_NAD = []
dataset_NAD_woPat4Pat11 = []
dataset_Ncontrol = []
dataset_control = []
dataset_AD = []
dataset_AD_woPat4_Pat11 = []
dataset_AD_woPat4_Pat11_woN = []

patients_AD = [1, 4, 10, 11, 12, 13, 14, 16, 17]
patients_control = [2, 3, 5, 6, 7, 8, 9, 15, 18, 19, 20, 21]
patients_AD_woPat4Pat11 = [1, 10, 12, 13, 14, 16, 17]

#Compare common dataset AD without patient N vs single patient NAD

R_AD = []

for j in range(9):
```

```python
        dataset_AD_woN.append(pd.read_excel(
            "Density_matrix_AD_woPat{}.xlsx".format(patients_AD[j])))
        dataset_NAD.append(pd.read_excel(
            "Density_matrix_patient{}.xlsx".format(patients_AD[j])))
        tempR = []
        tempR.append(pearson_coeff(dataset_AD_woN[j], dataset_NAD[j]))
        R_AD.append(tempR)


R_AD = np.array(R_AD)

np.savetxt('pearson_coefficient_AD.csv', R_AD, delimiter=',')


#Compare common dataset control without patient N vs single patient
Ncontrol

R_control = []

for k in range(12):

    dataset_control_woN.append(pd.read_excel(
        "Density_matrix_CR_woPat{}.xlsx".format(patients_control[k])))
    dataset_Ncontrol.append(pd.read_excel(
        "Density_matrix_patient{}.xlsx".format(patients_control[k])))

    tempR = []
    tempR.append(pearson_coeff(dataset_control_woN[k],
dataset_Ncontrol[k]))
    R_control.append(tempR)

R_control = np.array(R_control)

np.savetxt('pearson_coefficient_control.csv', R_control, delimiter=',')


#Compare common dataset control vs single patient NAD

R_NAD_vs_control = []

dataset_control.append(
    pd.read_excel("Density_matrix_CR.xlsx"))

for j in range(9):

    tempR = []
    tempR.append(pearson_coeff(
        dataset_control[0], dataset_NAD[j]))
    R_NAD_vs_control.append(tempR)

R_NAD_vs_control = np.array(R_NAD_vs_control)

np.savetxt('pearson_coefficient_NAD_vs_control.csv',
        R_NAD_vs_control, delimiter=',')



#Compare common dataset AD vs single patient N Control

R_Ncontrol_vs_AD = []

dataset_AD.append(pd.read_excel(
    "Density_matrix_AD.xlsx"))
```

```python
for i in range(12):

    tempR = []
    tempR.append(pearson_coeff(
        dataset_AD[0], dataset_Ncontrol[i]))
    R_Ncontrol_vs_AD.append(tempR)


R_Ncontrol_vs_AD = np.array(R_Ncontrol_vs_AD)

np.savetxt('pearson_coefficient_Ncontrol_vs_AD.csv',
           R_Ncontrol_vs_AD, delimiter=',')


#Compare common dataset AD without Pat 4 and Pat 11 vs single patient
Ncontrol

R_Ncontrol_vs_AD_woPat4Pat11 = []


dataset_AD_woPat4_Pat11.append(pd.read_excel(
    "Density_matrix_AD_woPat4_woPat11.xlsx"))

for i in range(12):

    tempR = []
    tempR.append(pearson_coeff(
        dataset_AD_woPat4_Pat11[0], dataset_Ncontrol[i]))
    R_Ncontrol_vs_AD_woPat4Pat11.append(tempR)

R_Ncontrol_vs_AD_woPat4Pat11 = np.array(R_Ncontrol_vs_AD_woPat4Pat11)

np.savetxt('pearson_coefficient_Ncontrol_vs_AD_woPat4Pat11.csv',
           R_Ncontrol_vs_AD_woPat4Pat11, delimiter=',')


#Compare common dataset AD without patient 4 and 11 vs single patient NAD

R_NAD_vs_AD_woPat4Pat11 = []


for l in range(7):

    dataset_AD_woPat4_Pat11_woN.append(pd.read_excel(
"Density_matrix_AD_woPat4_woPat11_woPat{}.xlsx".format(patients_AD_woPat4Pa
t11[l])))
    dataset_NAD_woPat4Pat11.append(pd.read_excel(
"Density_matrix_patient{}.xlsx".format(patients_AD_woPat4Pat11[l])))
    tempR = []
    tempR.append(pearson_coeff(
        dataset_AD_woPat4_Pat11_woN[l], dataset_NAD_woPat4Pat11[l]))
    R_NAD_vs_AD_woPat4Pat11.append(tempR)

R_NAD_vs_AD_woPat4Pat11 = np.array(R_NAD_vs_AD_woPat4Pat11)

np.savetxt('pearson_coefficient_NAD_vs_AD_woPat4Pat11.csv',
         R_NAD_vs_AD_woPat4Pat11, delimiter=',')
```

3.2 Paper II

Reference

Copyright statement

Contributions

I prepared the carbon nanodots, collected the AML and apheresate cytometry data, participated in the data analysis, wrote the initial draft and prepared the figures.

Check for updates

Uptake of carbon nanodots into human AML cells in comparison to primary hematopoietic cells†

Cathrin Nollmann,[a] Christian Wimmenauer,[a] Stefan Fasbender, [ID][a] Saskia Mayer,[b] Ron-Patrick Caddedu,[b] Paul Jäger,[b] Thomas Heinzel [ID] *[a] and Rainer Haas*[b]

Carbon nanodots (CNDs) comprise a class of next generation nanomaterials with a wide variety of potential applications. Here, we report on their uptake into primary hematopoietic cells from three normal donors and malignant cells from five patients with *de novo* acute myeloid leukemia (AML). A significant CND uptake was observed in all cell types of the normal and leukemic cells. Still, the uptake was significantly smaller for the CD34+ and CD33+ myeloid subsets of the malignant cell population as compared to the normal blood-derived CD34+ and CD33+ cells. For the T and B lymphoid cell populations as defined by CD3 and CD19 within the leukemic and normal samples a similar uptake was observed. The CNDs accumulate preferentially in small clusters in the periphery of the nucleus as already shown in previous studies for CD34+ progenitor/stem cells and human breast cancer cells. This particular subcellular localization could be useful for targeting the lysosomal compartment, which plays a pivotal role in the context of autophagy associated survival of AML cells. Our results demonstrate the usability of CNDs beyond their application for *in vitro* and *in vivo* fluorescence labeling or drug delivery into normal and malignant cells.

1. Introduction

Carbon nanodots, the family of carbon based, nanometer-sized particles which includes graphene quantum dots as well as small graphitic crystallites, have a large surface to volume ratio and excellent biocompatibility,[1,2] and studying new ways for their production is still an active field of research.[3–6] Since they also show fluorescence with advantageous properties in an aqueous environment, CNDs have been widely used in biomedical studies,[3,7–9] although the origin of the fluorescence is still under debate.[10–13] As far as living cells are concerned, CNDs enter the cytoplasm of many human cell lines as well as of primary human blood cells, without significant effects on the cell viability.[14–19] As for other nanoparticles, CNDs have been used in studies for cancer diagnosis or drug delivery.[20–26] In this context, the question arises whether the uptake of CNDs by malignant primary cells differs from that observed for healthy cells.

We therefore investigated the uptake of CNDs into leukemic cells that were freshly obtained from patients with *de novo* acute myeloid leukemia. In the majority of patients with this kind of leukemia, the pathological blasts resemble their normal

counterparts to some extent, in particular with respect to the expression of particular lineage- and differentiation associated surface molecules. Between those, CD34 and CD33 are prominent representatives reflecting an early stemness phenotype and myeloid differentiation, respectively. Our focus was on the aspect whether there is a differential uptake between primary human blood cells and leukemic cells, which could be of translational relevance.

2. Experimental section

2.1 Patients according to AML classification

The AML can be divided into different subtypes. The most commonly used classification schemes are the French–American–British (FAB)[27] and the World Health Organization (WHO)[28] system. The FAB classification is based on cytomorphological and cytochemical criteria, while the more recent WHO classification combines the FAB classification with immunological, cyto-as well as molecular genetic alterations. Table 1 lists the WHO and FAB classifications of the AML samples used in our study, as well as the percentage of blasts in the bone marrow and peripheral blood. The AML samples fall into different FAB categories permitting to some extent a subtype related assessment of the uptake.

2.2 Materials

Citric acid (ACS reagent, ≥99.5%), Diethylentriamine (DETA, 99%), Dulbecco's Phosphate Buffered Saline (DPBS), Lysis

[a]Condensed Matter Physics Laboratory, Heinrich-Heine-University, 40204 Düsseldorf, Germany. E-mail: Thomas.Heinzel@hhu.de

[b]Department of Haematology, Oncology and Clinical Immunology, Heinrich-Heine-University, 40204 Düsseldorf, Germany. E-mail: Haas.med@uni-duesseldorf.de

Table 1 WHO and FAB classifications of the five AML samples and the percentage of blasts in bone marrow (BM) and peripheral blood (PB)

Sample	WHO classification	FAB classification	Percentage of blasts
AML 1	AML with MDS-associated modifications	AML M0	PB: 39% BM: 32%
AML 2	AML with MDS-associated modifications	AML M2	PB: 65% BM: 69%
AML 3	Acute leukemia, assignment unclear	AML M4/5	PB: 10% BM: 40%
AML 4	AML without further cytometric or molecular genetic specification	AML M2	PB: 1% BM: 57%
AML 5	AML without further cytometric or molecular genetic specification	AML M1	PB: 77% BM: 85%

Buffer, Float-A-Lyzer dialysis devices (100–500 Da). Antibodies against CD45-PE-Cy7, CD34-PerCP-Cy5.5, CD33-PE, CD19-APC-R700 and CD3-APC-H7 were purchased from BD biosciences. Stem SPAN™ SFEM medium was bought at STEMCELL™ Technologies and microwave reaction vessels were obtained from CEM GmbH. NucBlue™ Live ReadyProbes™ Reagent (Invitrogen™), Poly-L-Lysine coated 8 well μ-slide was obtained from Ibidi.

2.3 Preparation and characterization of CNDs

The CNDs were prepared and characterized exactly as described in detail elsewhere.[18] In brief, fluorescent CNDs were synthesized according to the method of Qu *et al.*[29] with slight modifications. 210 mg citric acid and 340 mg Diethylentriamine (DETA) were mixed and heated to 180 °C for 150 s in a closed microwave reaction chamber (CEM Discover). The resulting viscous liquid was dissolved in 10 ml DI water. Citric acid, DETA and very small particles were removed by dialysis for 72 h, using a 100–500 Da membrane, with two water exchanges every 24 h. Afterwards, the CND solution was freeze-dried and dissolved for further use. The CNDs were characterized by scanning probe microscopy, transmission electron microscopy as well as by Raman- and XPS spectroscopy. These measurements indicate that the CNDs are composed of the mass fractions 40% C, 33% O, 19% N and 8% H. About 30% of the carbon bonds are C–C bonds, and the Raman spectra reveal their mixture of sp^2 with sp^3 hybridization. The average size of the CNDs was (3.3 ± 0.6) nm. Some particles showed a hexagonal crystal symmetry in the TEM with a lattice constant of approximately 220 pm, which is 10% smaller than that one of free-standing graphene. Hence, we conclude that the CNDs are particles with mixed crystal structures of graphene-, graphite- and possible diamond-type sections. Their photoluminescence properties are most relevant for the present study. They were measured using a Horiba FluoroMax®-4 spectrofluorometer, while the absorbance spectra were taken with an Agilent Cary 4000 spectrophotometer. The absorbance shows a peak around 360 nm wavelength of approximately 80 nm linewidth (full width at half maximum-FWHM). To avoid UV exposure of the cells, we excited the CNDs with light of $\lambda = 400$ nm, yielding a fluorescence spectrum centered at $\lambda = 460$ nm (FWHM ≈ 100 nm), which fits well into the V450 channel of the flow cytometer (see below), which we used for their detection. The quantum yield of the CNDs was 23%.[20] Since our particles do not show a fluorescence wavelength that differs from the values expected for size quantization and furthermore contain structural elements, like sp^3 hybridizations, that should be absent in graphene, we refer to

them as carbon nanodots. These nanoparticles are selected for the present study for several reasons. First of all, it has been shown earlier that they have a small toxicity and almost no influence on the gene expression of the exposed cells.[19] Second, they have a small mass and are thus expected to exert only a marginal influence on the dynamics of attached macromolecules of interest. Finally, they are relatively simple to prepare, show a competitive quantum yield and have a long shelf life of several months.

2.4 Collection of leukapheresis derived blood samples from normal donors

Primary hematopoietic cells were obtained from leukapheresis products (LP) of three healthy individuals who served as HLA-identical donors for an allogeneic blood stem cell transplantation using the granulocyte colony stimulating factor (G-CSF) at a dose of 480 μg per day over a period of five days. This increases the number of circulating human progenitor and stem cells (HSCs) – as characterised by the expression of CD34 on the cell surface – in the peripheral blood. This kind of LP samples are furthermore enriched with mononuclear blood cells (MNCs) including T and B cells, monocytes and CD34+ HSCs.[30–32] The samples also contain a significant percentage of granulocytes which are activated due to the exposure to G-CSF for five days.[33]

2.5 Cell preparation

Blood samples from LPs of three healthy donors and AML samples from five patients *de novo* AML were used for the *in vitro* studies. In order to remove the erythrocytes, the samples from the donors and from the patients were lysed with 50 ml ammonium chloride, once and twice, respectively. For all samples, the remaining leukocytes were resuspended in 50 ml PBS and centrifuged for 5 min at 300 g. Cell pellets were resuspended in 2 ml PBS. 3 ml per well of StemSpan™ Serum-Free Expansion Medium (SFEM) were dispensed in six well plates and appropriate amounts of cell suspension were added, resulting in a final concentration of 2×10^6 cells per ml.

2.6 Cultivation of the cells for CND uptake studies

CNDs were dissolved in PBS at concentrations of 20 mg ml^{-1}. The obtained solutions were sterile filtered. 75 μl of CND solution, corresponding to a concentration of 500 μg ml^{-1}, was added to the cell culture. The same amount of PBS without CNDs was added to the wells serving as negative control. The cells were cultivated in a Heracell™ 150i incubator in

Paper

a humidified atmosphere at 5% CO_2 and 37 °C. After 24 h, all samples were washed twice with PBS (centrifugation for 5 minutes at 300 g) and stained with antibodies as described in detail below.

2.7 Flow cytometry workflow

To study the differential uptake of CNDs in the various subpopulations of primary human blood cells and leukemic cells, monoclonal antibodies directed against lineage and differentiation specific antigens, *i.e.*, CD45-PE-Cy7, CD34-PerCP-Cy5.5, CD33-PE, CD19-APC-R700 and CD3-APC-H7 were used. The cells were incubated for 15 minutes in the dark with 2 µl diluted CD45 (1 : 10 with PBS), 2.5 µl CD34, 2.5 µl CD33, 1 µl CD19 and 1 µl CD3 per sample. Afterwards, they were washed with 2 ml PBS (centrifugation for 5 minutes at 300 g) and fixed with 200 µl 0.5% formaldehyde. To exclude results that do not originate from viable cells (*e.g.*, cell fragments or clumps), a gate was set in a forward *vs.* side scatter plot (FSC *vs.* SSC, see Fig. S1† in the ESI for examples). The FSC strength allows the discrimination of the cells by their size, while the SSC signal distinguishes the cell types by their granularity. Our gating strategy is exemplified for one AML sample and one donor sample each in Fig. 1. For an overall characterization of the sample composition, all viable cells from each sample are represented in a CD45 *vs.* SSC scatter plot (Fig. 1A and F). This also allows us to set blast gates for the AML samples later on.[34–36] Afterwards, the CD33$^+$ cells were gated out in a CD33 *vs.* CD45 plot, see Fig. 1B and G. Since the AML samples were extracted from the bone marrow (BM), the CD33$^+$ gate of the AML samples contains both myeloid progenitor cells and malignant blasts. The donor samples, on the other hand, are collected from the peripheral blood (PB), and hence the CD33$^+$ cells are mainly monocytes, mixed with some basophils (a subtype of granulocytes). In the next step, the stem and progenitor cells (CD45$^+$/CD34$^+$) were selected with a CD34 *vs.* CD45 plot as shown in Fig. 1C and H, respectively. The CD34$^+$ cells of the AML samples include the malignant blasts. Even though some of the CD33$^+$ cells of the donor samples have a high fluorescence intensity in the CD34 PerCp-Cy5.5 channel, they were not included in the CD34$^+$ gate since CD34$^+$/CD33$^+$ cells are absent in the peripheral blood. To distinguish between the lymphocyte subpopulations, CD3$^+$/ CD19$^-$ cells (T cells) as well as CD19$^+$/CD3$^-$ cells (B cells) were gated out in a CD45 *vs.* CD3 respectively CD45 *vs.* CD19 plot. In order to distinguish between the different types of blasts, further gates were set which are detailed in Section 2.7.

FACS analysis was performed using a BD FACSLyric™ flow cytometer. It is equipped with a 488 nm and a 640 nm laser to measure the fluorescence intensities in the PE, PerCP-Cy5.5, PE-Cy7, APC-R700 and APC-H7 channels and a laser with an excitation wavelength of 405 nm, allowing the measurement of the CND induced fluorescence in the V450 channel. For each sample, at least 50 000 events were recorded. The analysis was carried out using the BD FACSuite™ software.

The ratio of the mean V450 fluorescence intensity measured for the samples with CNDs to that one in the control samples was calculated, resulting in the uptake factor as the parameter,
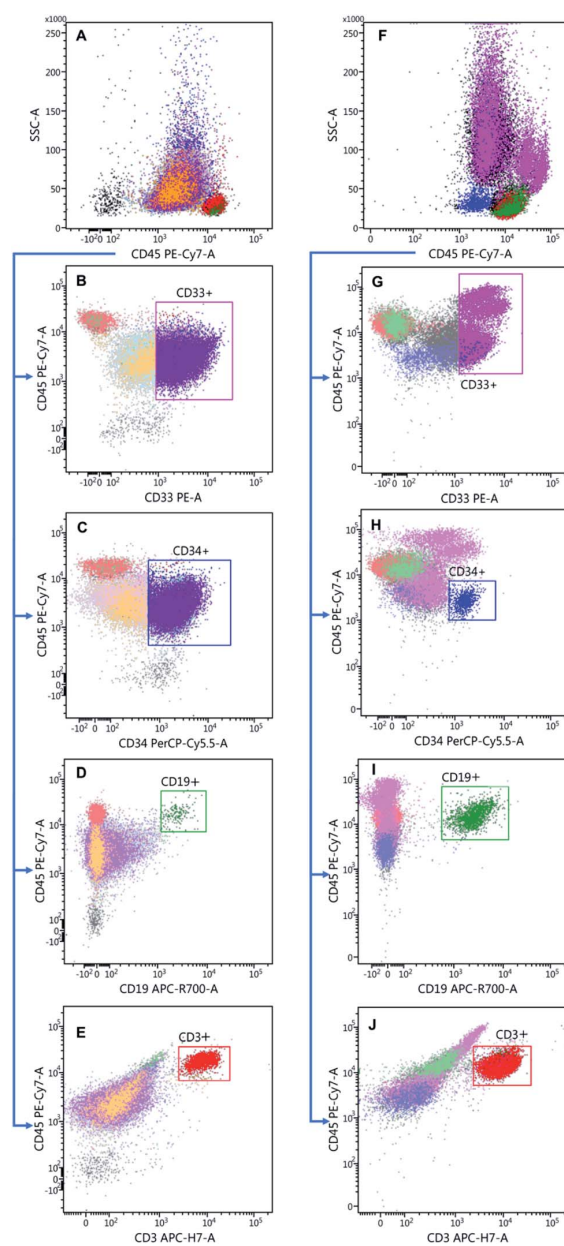

Fig. 1 Gating strategy of the AML samples in the left column and of the donor samples in the right column: scatter plots of the viable cells of AML 2 (A) and donor 3 (F) which are further differentiated using CD markers. Gating of the myeloid progenitor cells and CD33$^+$ malignant blasts (B) and the monocytes (G). (C) CD34$^+$ blasts respectively stem and progenitor cells gate (H). Differentiation between CD3$^-$/CD19$^+$ B cells and CD19$^-$/CD3$^+$ T cells of AML 2 (D and E) and of donor 3 (I and J).

which quantifies the cellular uptake of CNDs for each specific subpopulation. We thereby postulate that the CND fluorescence intensity represents a suitable parameter reflecting the local particle number. This implies that the intensity is not

RSC Advances

concentration-or pH-dependent. Concentration-dependent studies have shown that for the CND concentrations used here, the fluorescence intensity is linear as a function of the concentrations, while a significant pH dependence is observed only for unphysiologically low of high pH values (not shown).

2.8 Confocal fluorescence microscopy

The AML cell line HL-60 (passage number 18) was selected for the microscopy experiments. The cells were incubated in appropriate nutrition medium with a concentration of 500 μg ml^{-1} CNDs at 37 °C and 5% CO_2 for 48 h in a Poly-L-Lysine coated 8 well μ-slide, the nuclei were stained using NucBlue™ Live ReadyProbes™ Reagent (Invitrogen™) and the nutrition medium was exchanged for fresh medium without CNDs. The cells were imaged using a Zeiss LSM 710 confocal microscope evaluating the Hoechst 33342 channel (excitation 405 nm, emission 410–495 nm), the CND channel (excitation 488 nm, emission 495–530 nm) in framewise acquisition mode. A 63× oil objective with NA 1.40 was used.

2.9 t-SNE representation of the ensembles

Visualization of the multi-labelled cell ensembles in two dimensions by t-distributed stochastic neighbour embedding (t-SNE) has been carried out.[37,38] The FlowJo™ software has been used for this purpose. The perplexity was set to 30 and the number of iterations to 1000, respectively. The learning rate was automatically adjusted for every sample by FlowJo™ software. In our t-SNE plots, the cell subtypes appear in clusters, while a colour scale represents the CND fluorescence intensity. In order to attribute partially overlapping clusters to the corresponding cell types in the AML ensembles, overlays with the gated CD33$^+$, CD34$^+$, CD19$^+$ and CD3$^+$ populations are created, and the thereby identified populations are framed in the t-SNE plots accordingly.

2.10 Ethical statement

All experiments were performed in compliance with the relevant laws and institutional guidelines and have been approved by the ethical committee of the Heinrich Heine University (Study-no.: 2018-50_1). All donors had given their informed consent according to the guidelines of the ethical committee specified above.

3. Results and discussion

3.1 Subset analysis of AML samples

To investigate the CND uptake by the blasts depending on their maturity level, four categories were defined based on the expression level of CD33 and CD34. Undifferentiated blasts almost only express CD34 antigens on their surface, while CD33 gradually emerges at a later stage of maturation when CD34 is vanishing. To investigate the influence of the maturity level on the CND uptake, the blasts were gated within a CD45 vs. SSC plot for every AML sample (Fig. 2A and B). Afterwards, the blasts were classified according to the expression of CD33 and CD34 in
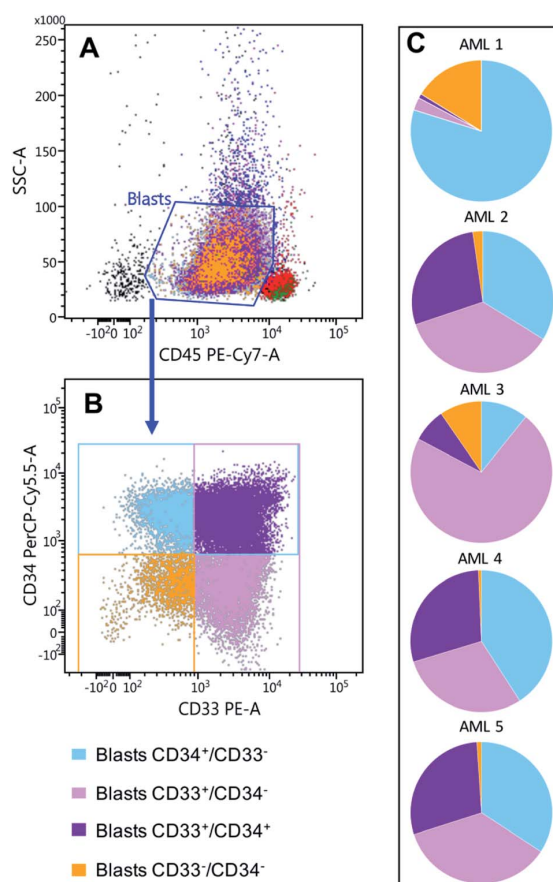
Fig. 2 AML samples characteristics: gating of the different blasts subsets, firstly all blasts were gated out in a CD45 vs. SSC plot (A). Secondly four categories were defined: CD33$^+$/CD34$^-$ blasts, CD34$^+$/CD33$^-$ blasts, blasts that were positive for both antibodies (CD33$^+$/CD34$^+$) and those that were not positive for either antibody (CD33$^-$/CD34$^-$) (B). (C) Distributions of the four blasts categories are shown for the five AML samples.

four subsets depending on whether they were positive or negative for the respective antigen.

The resulting distribution varied significantly between the different AML samples (Fig. 2C). While the CD33$^+$/CD34$^-$, CD34$^+$/CD33$^-$ or CD33$^+$/CD34$^+$ blasts are nearly equally distributed in AML 2, AML 4 and AML 5, CD34$^+$/CD33$^-$ blasts predominated in AML 1 and CD33$^+$/CD34$^-$ blasts in AML 3, reflecting the degree of relative maturity of the blasts within their pathological boundaries. The results correspond to some degree with the classifications of the AMLs (Table 1). AML 1 was classified as M0, i.e., a predominantly undifferentiated acute myeloblastic leukemia (CD34$^+$/CD33$^-$), whereas AML 2 and 4 belong to the M2 class (AML with maturation). AML 5 belongs to M1, an acute myeloblastic leukemia with some maturation, as indicated by the increase of the CD33$^+$/CD34$^+$ and CD33$^+$/CD34$^-$ blasts. Finally, in AML 3, the CD33+/CD34$^-$ blasts were dominant reflecting the M4/M5 classification (acute myelomonocytic/monocytic leukemia).

3.2 Cellular uptake of CNDs

We proceed by examining whether there is a differential uptake between primary human blood cells and the leukemic cells collected from the bone marrow, as well as between different subpopulations of the samples. In order to quantify and compare the uptake, we defined the uptake factor as the ratio of the mean fluorescence signal after CND exposure to that one of the negative control.

First, healthy and leukemic cells show an uptake of CNDs, reflected by a significant increase in signal intensity in the cells cultivated in the presence of CNDs as compared to the controls. The mean fluorescence signal in the V450 channel is increased by at least a factor of four. Representative examples are given in Fig. 3 for AML 2 (C) and for donor 3 (D). In more detail, the uptake factor for the $CD34^+$ cells (HSCs) from the donor samples is 1.7 – fold greater as compared to that one of the AML samples. The uptake factors of the $CD33^+$ populations differ even more between AML and donor samples, as the mean uptake factor for the donor samples is increased by 3.2. These findings indicate that the leukemic cells $CD33^+$ and $CD34^+$ leukemic cells have an apparently reduced ability to take up small compounds such as CNDs from the extracellular space.

Next, we have studied whether one of the four blast subpopulations (as obtained from the gating protocol shown in Fig. 2) shows a selective uptake behaviour. The resulting uptake factors do not show any differences between the various blast categories (see Fig. S2† of the ESI). This finding indicates that the degree of differentiation of the blasts is not related to the uptake capability of the CNDs, suggesting that a subset specific targeting of blasts without further modifications of the CNDs is not feasible.

Having a closer look on the intensity histogram of the $CD33^+$ population (Fig. 3(B) and (C)) it becomes apparent that the intensity of the donor $CD33^+$ cells show two peaks, one around 2×10^3 counts and a second one around 10^4 counts. This split was not found in the negative control, which represents the autofluorescence. Hence, the splitting indicates that two different $CD33^+$ cell types are present which differ with regard to their CND uptake behaviour, which can be related to $CD33^+$ monocytes only present in the LP products of the normal donors. This cell type is a prototype for a phagocytic cell implying that the CNDs are engulfed by vesicles related to the endolysosomal pathway, as suggested in earlier work.[19] On the other hand, the first peak is probably related to a small proportion of $CD33^+$ progenitor cells contained within the population of mobilized $CD34^+$ cells (see Fig. S3† in the ESI).

For the $CD19^+$ and $CD3^+$ populations, there was no significant difference between the uptake factors of AML and donor samples. Still, for both, AML as well as donor samples, the uptake factor for the $CD19^+$ subpopulation is significantly
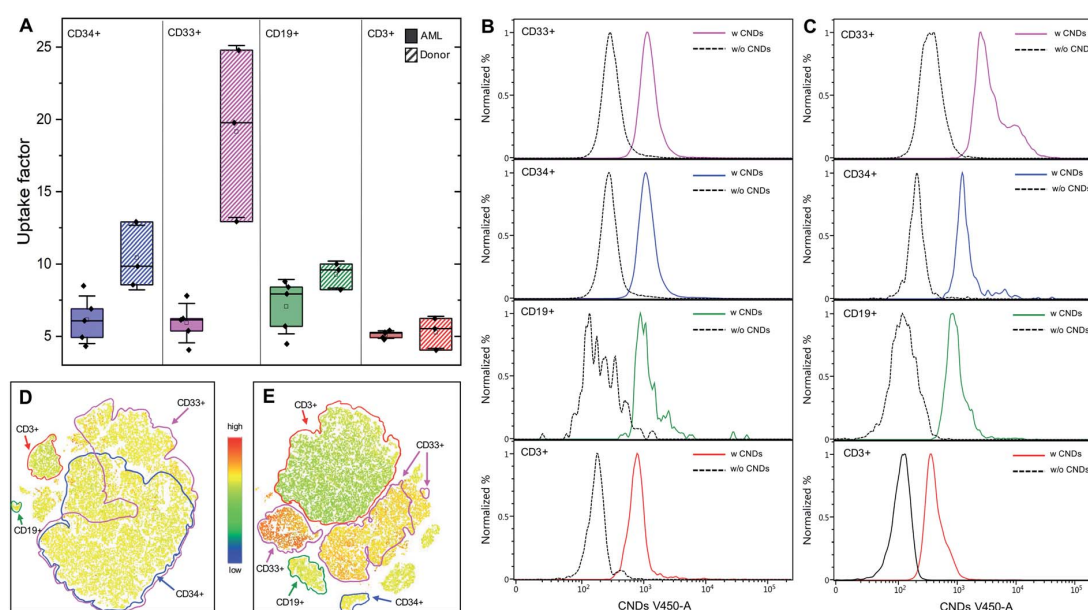

Fig. 3 (A) The sample-averaged uptake factors as determined for the four cell types and the corresponding statistical properties. The black squares denote the uptake factors of the individual samples, the horizontal bars are the median values and the unfilled squares are the mean values. The error bars indicate the standard deviation, and the lower and upper edge correspond to the first and the third quartile of the data. Please note that the $CD34^+$ population from the AML samples contains the $CD34^+/CD33^+$ and the $CD34^+/CD33^-$ cells. Likewise, the $CD33^+$ population is composed of the $CD34^+/CD33^+$ and the $CD34^-/CD33^+$ cells. Examples of the CND uptake by an AML (AML 2, (B)) and a healthy donor (donor 3, (C)) sample, as observed in the populations characterized by four CD antibodies. The intensity histograms are all normalized to a maximum value of 1 for better comparability. The t-SNE plots for the AML 2 and donor 3 samples are shown in (D) and (E), respectively as examples. Here, the colour scale quantifies the fluorescence intensity in the V450 A channel. The identified CD populations ($CD33^+$, $CD34^+$, $CD19^+$ and $CD3^+$) have been framed manually.

greater in comparison to the CD3$^+$ cells. For the AML and the normal donor samples, the mean uptake factors of the CD19$^+$ cells are about 1.38 and 1.76 times greater, respectively. A smaller uptake avidity of CD3$^+$ T and T helper cells compared to CD19$^+$ B cells from healthy donors was already observed in our previous study.[19] The difference between these two types of lymphoid cells may relate to the phagocytic ability of activated B cells.[39]

The t-SNE plots (Fig. 3(D) (for AML 2) and (E) (for donor 3)) permit a very informative illustration of our findings, as the fluorescence intensity of the CND signal in the V450 channel is color-coded. With regard to the AML samples, the relatively homogeneous intensity of colour across all subpopulations represents the similar uptake behaviour of the various cell types with a strong overlap of the CD33$^+$ and the CD34$^+$ cells. In contrast, the populations of the healthy samples are distinguishable in the t-SNE map, each of them characterized by a particular colour-coded uptake activity, which corresponds to the distinct peaks shown in Fig. 3C.

3.3 Subcellular distribution of CNDs in AML cells

In order to compare the intracellular distribution of the CNDs in AML cells, confocal microscopy images of HL-60 cells (corresponding to AML FAB M2 cells) were taken after 48 h of incubation with CNDs (Fig. 4). The CNDs accumulate preferentially in small clusters in the periphery of the nucleus. Based on the results of previous studies in HSCs and human breast cancer cells using a counterstaining method these clusters could be localized to lysosomes.[19,40] It is therefore conceivable that the CNDs following ingestion into the leukemic blasts are stored in the lysosomes, suggesting that the endolysosomal

pathway is also effective in AML cells. In the light of this finding the results of Folkerts et al. are interesting and of potential therapeutic relevance.[41] They could show decreased survival upon HCQ 20 μM hydroxychloroquine (HCQ) treatment for leukemic cell lines as well as primary sorted AML CD34$^+$ cells (n = 36) compared to normal bone marrow CD34$^+$ cells (N = 6; NBM CD34$^+$: 41.7% ± 7.1 vs. AML CD34$^+$: 21.3% ± 3.2, p = <0.05).

Microscopy images of the control samples are contained within in the ESI (see Fig. S4† and S5 in the ESI).

4. Conclusions

We compared the cellular uptake of small graphene quantum dots into normal blood cells with that into primary leukemic cells from patients with AML. Based on the intensity of the CNDs related autofluorescence recorded following a 24 h exposure time in an in vitro culture, a significantly smaller uptake was noted into leukemic cells compared to normal cells. This was true for both, the CD34$^+$ as well as CD33$^+$ subset. With regard to the uptake into lymphoid cells, a similar degree of uptake was observed for normal and leukemic cells, while a significant difference was only found between CD19$^+$ B cells and CD3$^+$ T cells irrespective of the sample source. This decreased differential uptake by the malignant cells studied here in comparison to their healthy counterparts forms a challenge for a selective addressing of those cells to which, e.g., a drug should be delivered. Suitable drug delivery systems based on our CNDs therefore may require some functionalization which increases the uptake by the target cells, like antigens or sugars, for example by preferential binding to the target cells. Alternatively, one might
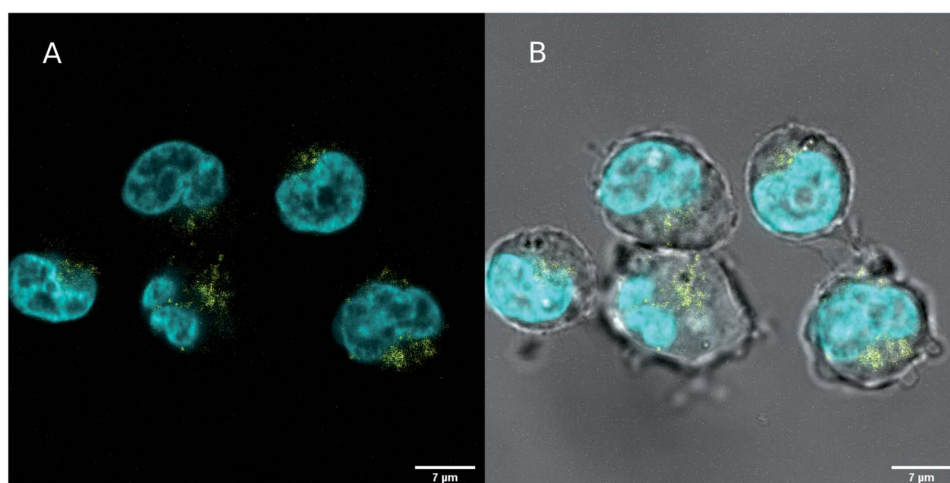

Fig. 4 Microscopy image of HL-60 cells taken 48 h after incubation with 500 μg ml^{-1} CNDs taken with a Zeiss LSM 710 confocal microscope (63× oil, NA 1.40). The nucleus was stained with Hoechst 33342, which was exited with a 405 nm UV diode laser and emission light was detected between 410 nm and 495 nm (cyan). The CNDs were exited with a 488 nm line from a multiline argon laser and the fluorescence was detected between 495 nm and 530 nm (yellow). The images have been acquired framewise. Image (A) displays only the fluorescence channels, while an overlay of the transmitted light from the CND channel and the fluorescence channels is shown in (B). Images of the control samples, taken with the same imaging parameters, are shown in the ESI.†

imagine selective inhibitor schemes where the CNDs protect the healthy cells from the impact of a drug. These issues are, however, beyond our scope here and will be the topic of future studies. It should be noted in this context that the reduced uptake of quantum dots in the malignant cell type studied here cannot be generalized to malignant cells of other organs, such as solid tumors of the breast or lung. The latter ones are epithelial in nature by their germline affiliation and thus may well differ with regard to their uptake properties from leukemia cells of mesenchymal origin.

Following their uptake, the CNDs reside in close proximity to endosomal–lysosomal machinery, which is involved in the uptake of extracellular particles *via* endocytosis. This subcellular location could be useful for therapeutic targeting involving the lysosomal compartment, which plays a pivotal role in the context of autophagy.[41] Watson and colleagues demonstrated the dual function of autophagy for the balance between cell death and cell survival.[42] They found that the complete blockade of autophagy induced the death of leukemic cells, while a reduction of this pathway increased their proliferation, which was associated with a significantly reduced latency of the disease. The dual role of autophagy for cancer progression and resistance is complex and therefore challenging when a targeted therapy is envisaged. Therefore, one direction of further work may be geared towards methods mediating a more specific leukemic uptake to assess functional effects on autophagy related processes in a dose-dependent manner.

List of abbreviations

AML	Acute myeloid leukemia
t-SNE	t-Distributed stochastic neighbour embedding
CNDs	Carbon nanodots
LP	Leukapheresis products

Author contributions

CN prepared the carbon nanodots, collected the AML and apheresate cytometry data, participated in the data analysis, wrote the initial draft and prepared the figures. CW took the confocal microscopy pictures, participated in the preparation and characterization of the CNDs and the data analysis. He furthermore initiated the t-SNE representation. SF contributed to the CND preparation setup and the flow cytometry. SM contributed to the recording of the cytometric data. RPD guided and supervised the cytometry and the gating of the cell ensembles. PJ collected the samples and evaluated the medical aspects of the AML data. TH co-designed the experiments, participated in the interpretation of the data and co-wrote the manuscript. RH initiated the study, interpreted the data and co-wrote the manuscript.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 Y. Yan, J. Gong, J. Chen, Z. Zeng, W. Huang, K. Pu, J. Liu and P. Chen, *Adv. Mater.*, 2019, **31**, e1808283.

2 H. Singh, S. Sreedharan, K. Tiwari, N. H. Green, C. Smythe, S. K. Pramanik, J. A. Thomas and A. Das, *Chem. Commun.*, 2019, **55**, 521.

3 K. Yadav, M. Das, N. Hassan, A. Mishra, J. Lahiri, A. K. Dubey, S. K. Yadav and A. S. Parmar, *RSC Adv.*, 2021, **11**, 976.

4 B. H. Lee, R. L. McKinney, M. T. Hasan and A. V. Naumov, *Materials*, 2021, **14**, 616.

5 M. J. Deka, P. Dutta, S. Sarnia, O. K. Medhi, N. C. Talukdar and D. Chowdhury, *Heliyon*, 2019, **5**, e01985.

6 P. Roy, P.-C. Chen, A. P. Periasamy, Y.-N. Chen and H.-T. Chang, *Mater. Today*, 2015, **18**, 447.

7 N. Panwar, A. M. Soehartono, K. K. Chan, S. Zeng, G. Xu, J. Qu, P. Coquet, K. Yong and X. Chen, *Chem. Rev.*, 2019, **119**, 9559.

8 J. Du, N. Xu, J. Fan, W. Sun and X. Peng, *Small*, 2019, **15**, 1805087.

9 W. Su, H. Wu, H. Xu, Y. Zhang, Y. Li, X. Li and L. Fan, *Mater. Chem. Front.*, 2020, **4**, 821.

10 R. Ludmerczki, S. Mura, C. M. Carbonaro, I. M. Mandity, M. Carraro, N. Senes, S. Garroni, G. Granozzi, L. Calvillo, S. Marras, L. Malfatti and P. Innocenzi, *Chem.–Eur. J.*, 2019, **25**, 11963.

11 J. Schneider, C. J. Reckmeier, Y. Xiong, M. von Seckendorff, A. S. Susha, P. Kasák and A. L. Rogach, *J. Phys. Chem. C*, 2017, **121**, 2014.

12 W. Kasprzyk, T. Świergosz, S. Bednarz, K. Walas, N. V. Bashmakova and D. Bogdał, *Nanoscale*, 2018, **10**, 13889.

13 W. Wang, B. Wang, H. Embrechts, C. Damm, A. Cadranel, V. Strauss, M. Distaso, V. Hinterberger, D. M. Guldi and W. Peukert, *RSC Adv.*, 2017, **7**, 24771.

14 L. Cao, X. Wang, M. J. Meziani, F. Lu, H. Wang, P. G. Luo, Y. Lin, B. A. Harruff, L. M. Veca, D. Murray, S. Xie and Y. Sun, *J. Am. Chem. Soc.*, 2007, **129**, 11318.

15 C. Liu, P. Zhang, F. Tian, W. Li, F. Lib and W. Liu, *J. Mater. Chem.*, 2011, **21**, 13163.

16 S. Zhu, J. Zhang, C. Qiao, S. Tang, Y. Li, W. Yuan, B. Li, L. Tian, F. Liu, R. Hu, H. Gao, H. Wei, H. Zhang, H. Sun and B. Yang, *Chem. Commun.*, 2011, **47**, 6858.

17 W. Shang, X. Zhang, M. Zhang, Z. Fan, Y. Sun, M. Han and L. Fan, *Nanoscale*, 2014, **6**, 5799.

18 S. Fasbender, S. Allani, C. Wimmenauer, R. P. Cadeddu, K. Raba, J. C. Fischer, B. Bulat, M. Luysberg, C. A. M. Seidel, T. Heinzel and R. Haas, *RSC Adv.*, 2017, **7**, 12208.

19 S. Fasbender, L. Zimmermann, R. P. Cadeddu, M. Luysberg, B. Moll, C. Janiak, T. Heinzel and R. Haas, *Sci. Rep.*, 2019, **9**, 12018.

20 C. Zhao, X. Song, Y. Liu, Y. Fu, L. Ye, N. Wang, F. Wang, L. Li, M. Mohammadniaei, M. Zhang, Q. Zhang and J. Liu, *J. Nanobiotechnol.*, 2020, **18**, 142.

21 Q. Liu and A. Atrens, *Corros. Rev.*, 2013, **31**, 85.

22 M. Zhang, L. Bai, W. Shang, W. Xie, H. Ma, Y. Fu, D. Fang, H. Sun, L. Fan, M. Han, C. Liu and S. Yang, *J. Mater. Chem.*, 2012, **22**, 7461.

23 I. Al-Ogaidi, H. Gou, Z. P. Aguilar, S. Guo, A. K. Melconian, A. K. A. Al-Kazaz, F. Meng and N. Wu, *Chem. Commun.*, 2014, **50**, 1344.

24 Z. L. Wu, M. X. Gao, T. T. Wang, X. Y. Wan, L. L. Zhenga and C. Z. Huang, *Nanoscale*, 2014, **6**, 3868.

25 X. Sui, C. Luo, C. Wang, F. Zhang, J. Zhang and S. Guo, *Nanomedicine*, 2016, **12**, 1997.

26 X. Gong, Q. Zhang, Y. Gao, S. Shuang, M. M. F. Choi and C. Dong, *ACS Appl. Mater. Interfaces*, 2016, **8**, 11288.

27 J. M. Bennett, D. Catovsky, M. T. Daniel, G. Flandrin, D. A. Galton, H. R. Gralnick and C. Sultan, *Ann. Intern. Med.*, 1985, **103**, 620.

28 S. H. Swerdlow, E. Campo, N. L. Harris, E. S. Jaffe, S. A. Pileri, H. Stein, J. Thiele and J. W. Vardiman, *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, WHO Press, Geneva, Switzerland, 2008.

29 D. Qu, M. Zheng, J. Li, Z. Xie and Z. Sun, *Light: Sci. Appl.*, 2015, **4**, e364.

30 R. Möhle, M. Pförsich, S. Fruehauf, B. Witt, A. Krämer and R. Haas, *Bone Marrow Transplant.*, 1994, **14**, 827.

31 K. Theilgaard-Mönch, K. Raaschou-Jensen, H. Palm, K. Schødt, C. Heilmann, L. Vindeløv, N. Jacobsen and E. Dickmeiss, *Bone Marrow Transplant.*, 2001, **28**, 1073.

32 M. D'Aveni, J. Rossignol, T. Coman, S. Sivakumaran, S. Henderson, T. Manzo, P. S. E. Sousa, J. Bruneau, G. Fouquet, F. Zavala, O. Alegria-Prvot, M. Garfa-Traor,

F. Suarez, H. Trebeden-Ngre, M. Mohty, C. L. Bennett, R. Chakraverty, O. Hermine and M. T. Rubio, *Sci. Transl. Med.*, 2015, **7**, 281ra42.

33 S. Murea, S. Fruehauf, W. J. Zeller and R. Haas, *J. Hematother.*, 1996, **5**, 351.

34 M. J. Borowitz, K. L. Guenther, K. E. Shults and G. T. Stelzer, *Am. J. Clin. Pathol.*, 1993, **100**, 534.

35 G. T. Stelzer, K. E. Shults and M. R. Loken, *Ann. N. Y. Acad. Sci.*, 1993, **677**, 265.

36 F. Lacombe, F. Durrieu, A. Briais, P. Dumain, F. Belloc, E. Bascans, J. Reiffers, M. R. Boisseau and P. Bernard, *Leukemia*, 1997, **11**, 1878.

37 L. van der Maaten and G. Hilton, *J. Mach. Learn. Res.*, 2008, **9**, 2579.

38 E. D. Amir, K. L. Davis, M. D. Tadmor, E. F. Simonds, J. H. Levine, S. C. Bendall, D. K. Shenfeld, S. Krishnaswamy, G. P. Nolan and D. Pe'er, *Nat. Biotechnol.*, 2013, **31**, 545.

39 Q. Zhu, M. Zhang, M. Shi, Y. Liu, Q. Zhao, W. Wang, G. Zhang, L. Yang, J. Zhi, L. Zhang, G. Hu, P. Chen, Y. Yang, W. Dai, T. Liu, Y. He, G. Feng and G. Zhao, *Immunobiology*, 2016, **221**, 558.

40 D. Kersting, S. Fasbender, R. Pilch, J. Kurth, A. Franken, M. Ludescher, J. Naskou, A. Hallenberger, C. von Gall, C. J. Mohr, R. Lukowski, K. Raba, S. Jaschinski, I. Esposito, J. C. Fischer, T. Fehm, D. Niederacher, H. Neubauer and T. Heinzel, *Nanotechnology*, 2019, **30**, 39.

41 H. Folkerts, B.-J. Wieringa, P. J. Coffer, J. J. Schuringa and E. Vellenga, *Blood*, 2015, **126**, 3831.

42 A. S. Watson, T. Riffelmacher, A. Stranks, O. Williams, J. de Boer, K. Cain, M. MacFarlane, J. McGouran, B. Kessler, S. Khandwala, O. Chowdhury, D. Puleston, K. Phadwal, M. Mortensen, D. Ferguson, E. Soilleux, P. Woll, S. E. W. Jacobsen and A. K. Simon, *Cell Death Discovery*, 2015, **1**, 15008.

Supplementary Information

Uptake of carbon nanodots into human AML cells in comparison to primary hematopoietic cells

Cathrin Nollmann[a], Christian Wimmenauer[a], Stefan Fasbender[a], Saskia Mayer[b], Ron-Patrick Caddedu[b], Paul Jäger[b], Thomas Heinzel[*a], and Rainer Haas[*b]

a. Condensed Matter Physics Laboratory, Heinrich-Heine-University, 40204 Düsseldorf, Germany.
E-Mail: Thomas.Heinzel@hhu.de
b. Department of Haematology, Oncology and Clinical Immunology, Heinrich-Heine-University, 40204 Düsseldorf, Germany.
E-Mail: Haas.med@uni-duesseldorf.de
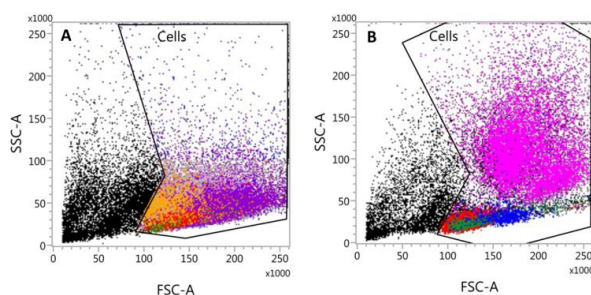
Fig. S1: Examples of FSC vs. SSC plots by which the cell population was selected via suitable gates. Shown are AML 2 in (A) and donor 3 in (B). The colours represent the rear projections of further gates that have been set, namely CD33+ (pink), CD34+ (blue), CD19+ (green), CD3+ (red), CD33+/CD34+ (purple) and CD33-/CD34- (orange).

Tab. S1: Percentage distribution of the CD33+, CD34+, CD19+ and CD3+ populations for all samples.

sample	CD33+	CD34+	CD19+	CD3+
AML 1	7,96%	14,40%	0,25%	0,84%
AML 2	77,01%	72,55%	0,55%	4,87%
AML 3	88,16%	3,82%	1,69%	6,33%
AML 4	42,11%	35,50%	1,11%	31,30%
AML 5	84,82%	77,05%	0,08%	1,21%
donor 1	33,14%	1,89%	2,48%	43,87%
donor 2	43,94%	3,56%	8,75%	36,93%
donor 3	24,15%	2,13%	4,64%	47,12%
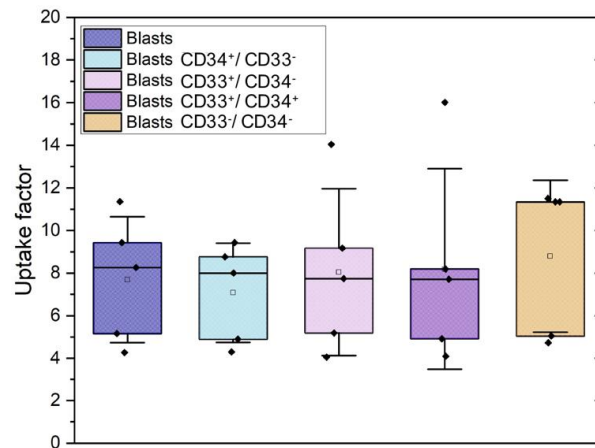
Fig. S2: Uptake factors for all blasts and four subsets which were defined by means of the CD34 and CD33 markers.
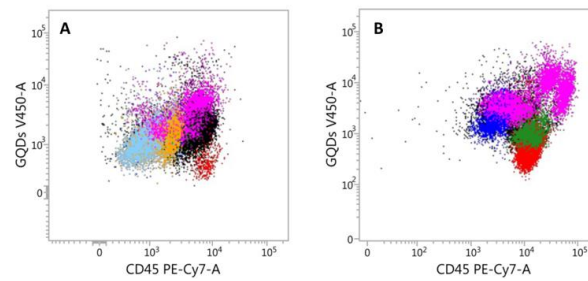
Fig. S3: CD45 vs. V450 plot, shown exemplarily for AML 1 (A) and donor 3 (B). The colours are rear projections of other set gates CD33$^+$ (pink), CD34$^+$ (blue), CD19$^+$ (green), CD3$^+$ (red), CD34$^+$/CD33$^-$ (light blue) and CD33$^-$/CD34$^-$ (orange).
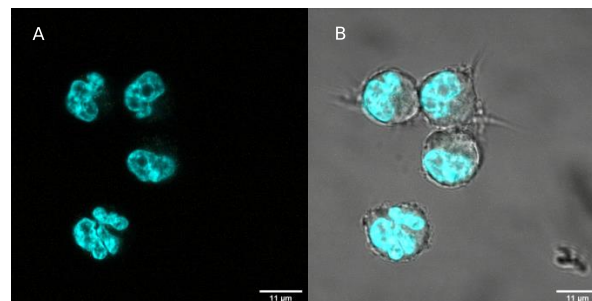
Fig. S4: Control samples (HL-60 cells) without CNDs. The nuclei are stained with Hoechst 33342.
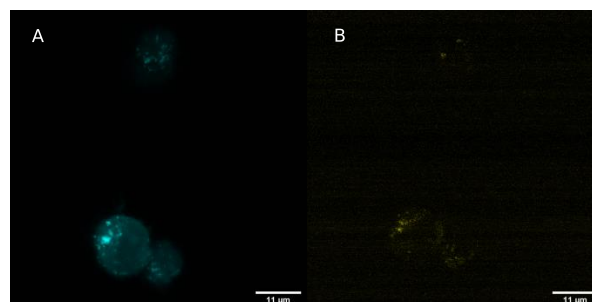
Fig. S5: Control sample of CND-exposed HL-60 cells with the nuclei unstained. The CNDs show fluorescence in the blue (a) and in the yellow (B) channel.

4. Further experiments on the uptake of CNDs

As an additional parameter in the AI-assisted evaluation of flow cytometry data from healthy donors compared to AML patients, the uptake of CNDs was investigated in *Paper II*. This investigation revealed differences in the uptake between diseased and healthy individuals. This observation gave rise to the question of whether a more selective uptake could be achieved by functionalizing the CNDs, which was investigated in *Paper III* by glycofunctionalizing the CNDs with various sugar monomers and glycooligomers. No differential uptake could be observed between the different cell types, whereas the various nanoparticle conjugates showed different uptake rates between the investigated cell lines. The investigations in *Paper II* also showed that the CNDs accumulate primarily in the lysosomes. This led to the question of which influence the CNDs have on the lysosomes and the associated process of autophagy and whether it is possible to deliver drugs specifically to the lysosomes using the CNDs as carriers. This question was investigated in *Paper IV*, where the influence of pristine CNDs and CNDs modified with bPEI was examined. The bPEI was coupled to the CNDs as an example of a drug. It was shown that the CNDs are inert as carriers and have no significant influence on lysosomal processes and autophagy characterized by cathepsins B and L, the autophagic markers SQSTM1/p62 and LC3, and proteome analysis. The bPEI was successfully transported into the lysosomes by the CNDs and led to an increased accumulation of bPEI-CNDs in the lysosomes compared to bPEI or CNDs alone.

4.1 Paper III

Reference

S. Üclü, C. Wimmenauer, C. Nollmann, A. Liu, S. A. Hill, N. L. Snyder, T. Heinzel and L. Hartmann (2023). Sweet carbon dots: synthesis of sugar-functionalized carbon nanoparticles and their cellular uptake. *To be resubmitted.*

Copyright statement

This manuscript is a preprint we are currently planning to resubmit.

Contributions

I prepared the samples for flow cytometry together with CW and carried out the measurements.

Sweet carbon dots: synthesis of sugar functionalized carbon nanoparticles and their cellular uptake

Serap Üclü, [‡a] Christian Wimmenauer, [‡b] Cathrin Nollmann,[b] Andrea Liu,[c] Stephen A. Hill,[a] Nicole L. Snyder,[c] Thomas Heinzel[*b] and Laura Hartmann[*a]

a. Department for Organic Chemistry and Macromolecular Chemistry, Heinrich Heine University Düsseldorf, Universitätsstraße 1, Düsseldorf 40225, Germany. E-mail: laura.hartmann@hhu.de

b. Institute for Experimental Condensed Matter Physics, Heinrich Heine University Düsseldorf, Universitätsstraße 1, Düsseldorf 40225,Germany. E-mail: thomas.heinzel@hhu.de

c. Department of Chemistry Davidson College 102 North Main Street, Davidson, NC 28035 USA E-mail: nisnyder@davidson.edu

Carbon nanoparticles are coupled covalently to various monosaccharides as well as glycooligomers and studied for their interaction with different tumor cell lines overexpressing carbohydrate-recognizing lectin receptors. Monosaccharide-conjugated carbon nanoparticles show an enhanced uptake rate as compared to both pristine as well as glycooligomer-coated particles. However, coating with mannose, galactose, N-acetylglucosamine or glycooligomers derived thereof does not influence the intracellular distribution of the nanoparticles showing that they are preferably stored in the endolysosomal pathway.

Functionalized nanoparticles and their application in biomedical research have developed into a mature scientific field over the last two decades.[1-3] Multicolor fluorescence spectroscopy based on quantum dots is used routinely within various diagnostic medical protocols.[4-6] Carbon nanoparticles (CNPs) are a promising complement to conventional semiconducting quantum dots in these respects, due to their low mass, intrinsic water solubility, easy functionalization and low toxicity.[7,8] They have been used as intracellular pH meter,[9] selectivity enhancers during cisplatin[10] or doxorubicin[11] uptake, or in confocal fluorescence microscopy[12-14], to name just a few examples. One important advancement in the development of nanomaterials for biomedical applications is their selectivity, e.g., for cancer vs. healthy cells.[15,17] Such selectivity can enable more efficient drug delivery and thus reduced side effects in anti-cancer therapy.[18-20] Interactions between carbohydrates and lectins as receptors have been shown to allow for such selectivity and have been successfully applied to derive various glycofunctionalized nanomaterials for use in biomedicine.[21-22] Examples of glycofunctionalization for cell specific targeting include galactose (Gal) ligands addressing the asialoglycoprotein receptor, sialic acid (Neu5Ac) ligands targeting Siglecs,[23] and hyaluronic acid (HA) functionalization for binding to CD44 receptors,[24]; each receptor is overexpressed in a specific cell type. Another carbohydrate recognizing C-type lectin receptor (CLR) with relevance for targeting cancer cells is the Mannose Receptor (MR). MR recognizes mannose (man) , fucose (fuc) and *N*-acetylglucosamine (GlcNAc) in a calcium-dependent manner.[25,26] Different types of cancer cells are known to overexpress MR,[27] and it has been shown that by targeting the MR e.g., through binding with its carbohydrate ligands, cancer cells can be distinguished from healthy cells.[28] In this study we aim to combine the advantages of CNPs and glycofunctionalization, with a special focus on targeting MR overexpressing cancer cells, and investigate the cellular uptake of glyco-CNPs in dependence of the type of glycofunctionalization using used and comparing different monosaccharides as well as glycooligomers.
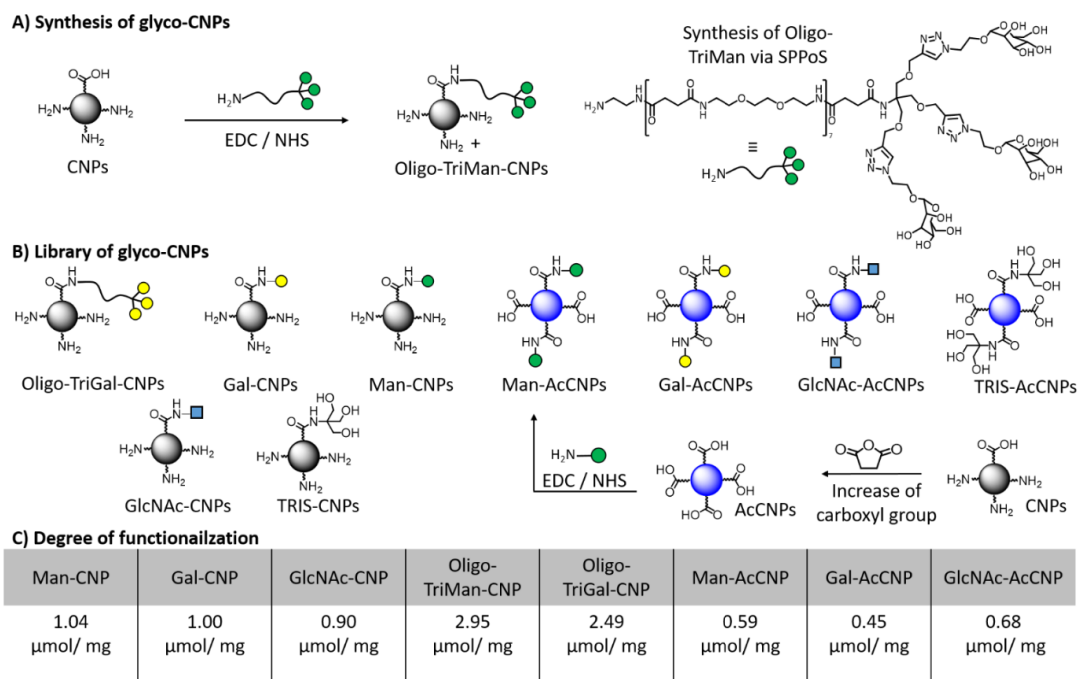
Figure 1: A) Synthesis of glyco-CNPs: SPPoS of Oligo-TriMan functionalization of CNPs via activation the carboxyl groups by EDC/NHS. B) Library of synthesized glyco-CNPs including the extension of carboxyl group of CNPs through succinic anhydride to gain higher level of functionalization degree (AcCNPs) and their glyco-conjugated derivatives. C) Functionalization degrees of glyco-CNPs.

Here, we present the preparation and characterization of the glyco-CNPs (Figure 1A), followed by uptake studies using three ubiquitous cell lines, which were exposed to a set of CNPs coupled to different monosaccharides as well as glycooligomers. We opted to use CNPs prepared by microwave assisted pyrolysis from citric acid and diethylenetriamine as described in detail the ESI, leading to free acid- and primary amine groups at the surface of the CNPs. A second fraction of the particles was first coupled to succinic anhydride (AcCNPs) thereby reacting with amine groups also present on the CNP surface and increase the degree of carboxylic groups, which are required for conjugation of the carbohydrate derivates and glycooligomers. (Figure 1B) In addition to acceptable fluorescence properties, these CNPs are very small with diameters of roughly 3 nm, and extremely stable. Furthermore, they are known to have only a marginal toxicity, even on the gene expression level.[29] For glycofunctionalization, a set of monosaccharides were used: Man and GlcNAc as known MR ligands, and Gal as a negative control.

Additionally, glycooligomers presenting multiple copies of Man or Gal were synthesized and applied for the functionalization of the CNPs. Cancer cells such as MDA-MB-231 breast cancer can be targeted by receptor-mediated endocytosis through overexpressed MR. [30,31] The multivalent presentation of Man is expected to increase binding and thus cell uptake, through the statistical possibility of a Man molecule encountering a CRD, resulting in binding enhancement.[32] Thus, glycooligomers were synthesized by previously established solid-phase polymer synthesis (SPPoS, see ESI).[33,34] In brief, stepwise conjugation of tailor-made building blocks using standard Fmoc-peptide coupling protocols on resin provided access to sequence-defined oligo(amidoamines) with alkyne side chains, which were used for further conjugation of carbohydrate-azide derivatives via a copper-mediated click reaction (CuAAc).[35] Glycooligomers used for subsequent conjugation to NPs carry a long hydrophilic chain and three terminal Man ligands. For later use as negative control, glycooligomers presenting Gal were synthesized accordingly.
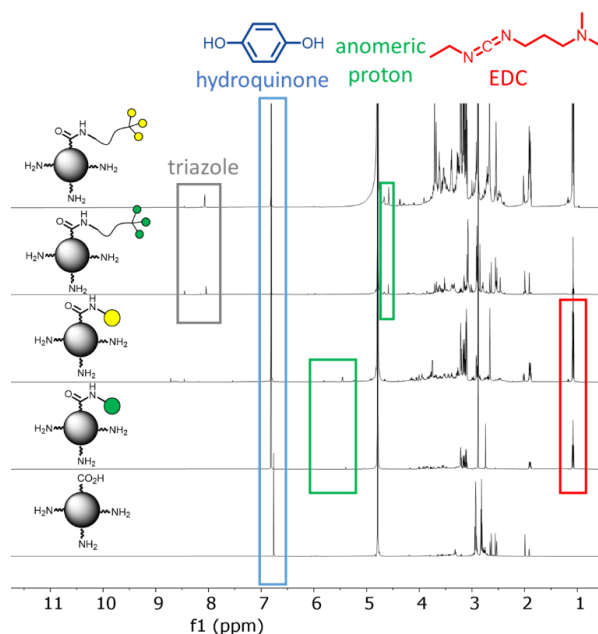
Figure 2: ^{1}H NMR spectra of pristine and functionalized CNPs. Hydroquinone was used as internal standard (blue frame). Green frame: anomeric proton signal of Man/Gal. Grey frame: signal from the protons of the triazole ring of the glycooligomers. The EDC signal is framed in red.

Conjugation of both, monosaccharides and glycooligomers to the CNPs were achieved by activation of carboxylic groups on the CNPs via 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and *N*-hydroxysuccinimide (NHS) followed by subsequent reaction with free amine groups either from amino-functionalized monosaccharides (see ESI for synthesis and analytical data of the amino sugars) or on the C-terminal position of the oligomer. The hydrophilic chain of the glycooligomer provides flexibility and may potentially increase the accessibility of the carbohydrate ligands in binding to cell surface receptors (see the ESI for further details on the synthesis and analytical data of the products).

The degree of functionalization of the glycol-CNPs was studied by ^{1}H NMR, see Figure 2. Hydroquinone was used as an internal standard and compared to either the anomeric protons of the sugar (green frame) or the protons of the triazole ring of the glycooligomers (grey frame), which provided the overall sugar concentration. Successful glycofunctionalization was demonstrated for both AcCNPs and CNPs, but with different degrees of functionalization. While the AcCNPs have a larger number of acid groups indicating a higher degree of functionalization, they have a lower sugar concentration as compared to the glycofunctionalized CNP system, see Figure 1 C. We attribute this to steric effects, since succinic anhydride functionalization limits the accessibility of the acid groups. In the CNP system, on the other hand, the acid groups have more space on the CNP surface, which results in better accessibility for glycofunctionalization. Additionally, the CNP system was used for functionalization with glycooligomers, confirming and quantifying the degree of functionalization via NMR (see Figure 2). In the NMR spectra, we also identified an EDC contamination (red frame in Figure 2) for the monosaccharide functionalized CNPs. Since we were unable to remove this contamination with extensive washing protocols, this is likely either covalently attached to the surface of the CNPs or located in between the carbohydrates through ionic interactions. Since the EDC remained present also after dialysis against high ionic buffer (PBS), covalent binding strikes us as more plausible. This contamination is not seen in the CNPs functionalized with glyco-oligomers, which could be because the flexible chain covers a larger surface area on the CNPs, thus preventing EDC binding to the surface. Diffusion ordered spectroscopy (DOSY) measurements were performed to evaluate the change of the
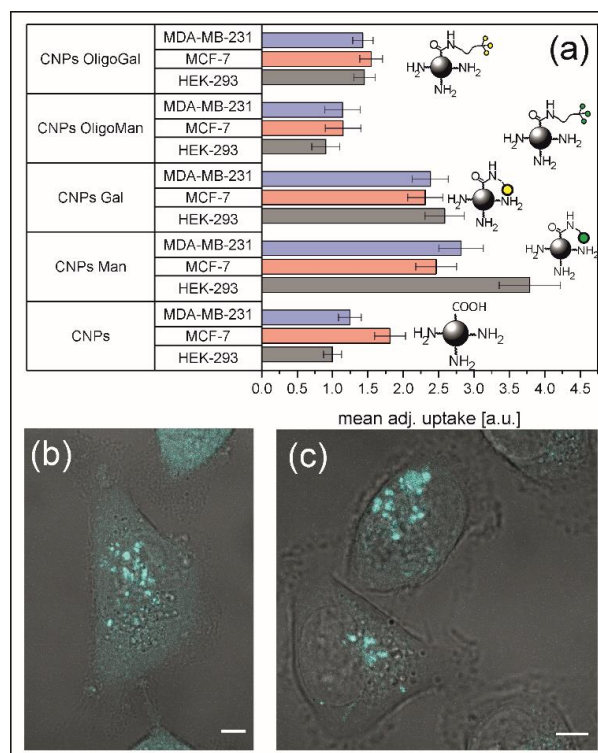
Figure 3: (a) CNP uptake as determined from flow cytometry measurements. Confocal microscopy on MDA-MB-231 cells after uptake of pristine (b) and mannose-conjugated (c), CNPs. The size of the scale bars is 5 μm.

hydrodynamic radii before and after the functionalization. All diffusion coefficients lie in the range [4 – 6] x 10^{-10} m^2 s^{-1},corresponding to hydrodynamic radii around 0.4 nm, almost independent of the functionalization (see Table 1 in the ESI). Fluorescence spectroscopy reveals that all functionalized CNPs show a weak but noticeable redshift of the fluorescence maximum, while the shape of the spectrum remains approximately constant (see ESI Table 2).

With this set of glyco-CNPs in hand, we then performed flow cytometry, confocal microscopy, and XTT assays to determine their cellular uptake rates, subcellular distribution, and cell toxicity, respectively. Two breast cancer cell lines MCF-7 and MDA-MB-231 as well as the embryonic kidney cell line HEK-293 were employed in these experiments based on their expression of MRs: HEK-293 cells were shown to present few MR in previous work, while MDA-MB-231 are known to possess a high expression of MRs.[36] MCF-7 cells were selected due to their high uptake of CNPs shown in previous studies.[13] To quantify cellular uptake, measured in terms of the total mass of nanoparticles per cell, the fluorescence intensity of an event detected in flow cytometry was corrected with the brightness of each particle species (see ESI Table 2). Here, the quantum yield under an excitation wavelength of 360 nm was determined using Coumarin 1 with a quantum yield of 0.5 as a reference. The excitation coefficient is obtained from a fit of Lambert-Beer's law to the concentration dependence of the absorption. We define the brightness of a particle type as the product of the quantum yield and the extinction coefficient. Increased uptake rates of CNPs functionalized with monosaccharides compared to non-functionalized CNPs were observed as shown in Figure 3 (a). This effect is independent of the cell type and therefore of the receptor expression. This was confirmed by additional receptor-specific studies with HepG2 cells, which are widely used for drug development and toxicity testing and overexpress the asialoglycoprotein receptor (ASGPR)[37] and MR[38]. Here, the specific uptake of Gal-CNPs targeting ASGPR of HepG2 cells were investigated, and showed the same uptake rate as the negative control cell line HEK293 (see the ESI). In addition, the influence of glycofunctionalization on the subcellular distribution of the glyco-CNPs was investigated, see Figure 3 (b), (c). Confocal microscopy revealed that

they accumulate in small sub-spaces inside the cell, which most likely correspond to vesicles in the endo-lysosomal pathway, an observation that has been confirmed via other studies.[13] Pristine and glycofunctionalized CNPs were observed to share this characteristic subcellular distribution 24 h after the incubation. Co-staining of the lysosomes with CellLight Lysosomes-RFP (Invitrogen) reveals a high correlation of the distribution of the lysosome resident protein LAMP1 and the CNPs (See ESI, Figure S36 + S37), indicating that all the CNPs used in our studies tend to accumulate in the lysosomes. Furthermore, glycofunctionalization does not lead to a decrease in cell viability 24 h post incubation, as measured via XTT assay (see the ESI Figure S38 for details). Surprisingly, CNPs functionalized with glycooligomers show uptake rates similar to the prisitine CNPs. We tentatively explain this uptake behaviour using the following, qualitative picture: During endocytosis, both pristine and oligomer-conjugated CNPs are taken up from the encapsulated extracellular solution; the uptake rate is proportional to the density of the CNPs in solution. In addition to this process, monosaccharide-conjugated CNPs may adhere to the outer membrane surface. This picture is commensurate with a quantitative estimation (see ESI). Since the uptake rates are independent of the cell type and hence of the densities of the sugar receptors at the cell surface, we conclude that interactions with receptors do not play a major role in CNP uptake. Rather, we conjecture that the increased uptake rate of monosaccharide-conjugated CNPs may originate from non-specific, physical interactions. The fact that glycoconjugation of the CNPs does not influence the viability of the cells supports the picture that the nanoparticles are captured in the endolysosomal pathway, where they exert only a marginal influence on cellular metabolism.

In conclusion, we have presented protocols for the glycofunctionalization of carbon nanoparticles with different monosaccharides, as well as glycooligomers. NMR studies have been used to verify the nanoparticle-sugar coupling. Uptake studies with different cancer cell lines reveal that the uptake rate does not correlate with the cell-specific expression of lectin receptors nor the type of carbohydrate. Rather, CNPs functionalized with sugar monomers show an approximately two- to threefold increase of the uptake rate in comparison to the pristine nanoparticles, which is similar to the uptake rate for oligomer-functionalized CNPs. This behaviour is tentatively explained using a geometric picture where the uptake takes place in an unspecific combination of CNPs in solution and those which adhere to the cell membrane. Confocal microscopy shows that glycofunctionalization does not modify the intracellular distribution, where the nanoparticles are preferably stored in the endolysosomal pathway. While carbohydrate functionalization has not led to an increase in cell selectivity, cell uptake increased with monosaccharide functionalization, making these glyco-CNPs potentially interesting for cell imaging or drug delivery applications. To achieve a more selective uptake, future studies could explore the effects of using more complex oligosaccharide ligands or non-carbohydrate ligands such as antibodies.

There are no conflicts to declare.

References

1 X. Michalet et al., Science, 2005, **307**, 538.
2 I. L. Medintz et al., Nature Materials, 2005, **4**, 435.
3 W. R. Algar et al., Chem. Rev., 2021, **121**, 9243.
4 J. Kim et al., Chem. Soc. Rev., 2009, **38**, 372-390.
5 L.-L. Chen et al., Small, 2022, **18**, 2104567.
6 R. Thangam et al., Nanomaterials, 2021, **12**, 18.
7 J. Shen et al., Chem. Commun., 2012, **48**, 3686.
8 L. Cao et al., J. Am. Chem. Soc., 2007, **129**, 11318.

9 Wu, Z. L. et al., Nanoscale **6**, 3868 (2014).

10 Sui, X. et al., Nanomedicine: Nanotech., Biol. Med., 2016, **12**,1997.

11 Gong, X. et al., ACS Appl. Mater. Interfaces, 2016, **8**, 11288.

12 J. Du, N. Xu, J. Fan, W. Sun and X. Peng, Small, 2019, **15**, 1805087.

13 D. Kersting et al., Nanotechnology, 2019, **30**, 39.

14 N. Panwar et al., Chem. Rev., 2019, **119**, 9559.

15 J. Chen et al., Adv. Sci. (Weinh.), 2016, **3**, 1600270.

16 Yim et al., Chem. Commun., 2021, **57**, 9820.

17 J. Sun et al., Chem.Commun., 2021, **57**, 1989.

18 C. Wang et al., J. Colloid Interface Sci., 2018, **516**, 332.

19 D. Rosenblum et al., Nat. Commun., 2018, **9**, 1410.

20 R. Lima-Sousa et al., Carbohydrate Polymers, 2018, **200**, 93.

21 N. Kong et al., ACS Appl. Bio Mater.,2019, **2**, 284.

22 D. Sharma et al., Colloid Polymer Science, 2020, **298**, 365.

23 M.P. Lenza et al., Cells, 2020, **9**, 2691.

24 J.M. Rios de la Rosa et al., Adv. Healthcare Mater., 2019, **8**, 1901182.

25 M. E. Taylor et al., J. Biol. Chem., 1992, 267, 1719.

26 P. R. Taylor et al., Immunol, 2005, 26, 104.

27 E. D. Vedove et al., Adv Healthc. Mater., 2018, **7**, 1701398.

28 L. Li et al., Nanomedicine, 2018, **14**, 2355.

29 S. Fasbender et al., Sci. Rep., 2019, **9**, 12018.

30 D. Brevet et al., Chem. Commun. 2009, **12**, 1475.

31 Y.-H. Tang et al., Biosens. Bioelectron. 2018, **116**, 100.

32 L. Raviv et al., Mol. Pharmaceutics, 2015, **12**, 453.

33 F. Wojcik et al., *Chemistry - A European Journal* **2013**, *19*, 3090-3098.

34 Ponader et al., J. *Am. Chem. Soc.,* **2014**, *136,* 2008−2016.

35 S. Hill and M. C. Galan, Beilstein J. Org. Chem., 2017, **13**, 1136.

36 Ye et al., Chem. Eur. J., 2016, **22**, 15216–15221.

37 Y. Li et al., Curr. Drug Delivery 2008, **5**, 299.

38 K. Staines et al., PLoS One 2014, **9**, e110330.

4.2 Paper IV

Reference

C. Sprengel, C. David, L. Berning, C. Nollmann, T. Lenz, K. Stühler, T. Heinzel and B. Stork (2024). Lysosomal activity in response to incubation of pristine and functionalized carbon nanodots. *Submitted.*

Copyright statement

This manuscript is a submitted preprint.

Contributions

I participated in conceiving the experiments, in analyzing and discussing the data. I assisted in the carbon nanodot functionalization and sample preparation for cell experiments as well as performed some confocal microscopy measurements. I contributed to writing, reviewing and editing of the final manuscript.

Lysosomal activity in response to incubation of pristine and functionalized carbon nanodots

Carla Sprengel ◉,[1] Céline David ◉,[2] Lena Berning ◉,[2] Cathrin Nollmann ◉,[1]
Thomas Lenz ◉,[3] Kai Stühler ◉,[3, 4] Thomas Heinzel ◉,[1, *] and Björn Stork ◉[2, †]

[1] *Solid State Physics Laboratory, Heinrich Heine University Düsseldorf, 40204 Düsseldorf, Germany*
[2] *Institute of Molecular Medicine I, Medical Faculty and University Hospital Düsseldorf,*
Heinrich Heine University, 40225 Düsseldorf, Germany
[3] *Molecular Proteomics Laboratory, Biological Medical Research Center,*
Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[4] *Institute of Molecular Medicine I, Proteome Research,*
Medical Faculty and University Hospital Düsseldorf,
Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
(Dated: June 22, 2024)

We present functional studies of lysosomes in human cells after uptake of carbon nanodots (CNDs). Even under high CND concentrations, the lysosomal functionality, as characterized via cathepsins B and L as well as the autophagic markers SQSTM1/p62 and LC3B, is maintained. Furthermore, branched polyethylenimine (bPEI) molecules have been coupled to the CNDs as a model functionalization or example of a drug. We observe that the bPEI-CND conjugates accumulate to a higher degree in the lysosomes as compared to bPEI or CND alone. Here, changes in the lysosomal size and function are observed, which can be explained exclusively by the bPEI. It is concluded that CNDs are highly efficient and inert carriers for functional molecules into lysosomes as target, with the added value that lysosomal escape is suppressed, thereby avoiding unintended side effects in other cellular compartments.

INTRODUCTION

Carbon nanodots (CNDs) are promising fluorophores for cellular imaging, since they combine a low mass with relatively low toxicity [1, 2], high water solubility [3, 4] and easy functionalizability [5]. In recent years, several studies have addressed their uptake by cells, their intracellular distribution as well as the metabolic and genetic response of the cells to their presence even at very high concentrations. Remarkable results have been reported, like for example, their suitability as intracellular pH sensors [4, 6], drug delivery options [7–9], application in photosensitizing experiments [10, 11] or the observation that the gene expression of the cells remains essentially unaffected in the presence of even approximately one hundred million CNDs per cell [2].

This raises expectations regarding their application in drug delivery, with hopes that their presence neither alters drug effects nor causes side effects by themselves. In this context, the distribution pathway of the CNDs in the cell after exposure is relevant. It has been shown that unmodified CNDs are primarily taken up via the endolysosomal pathway and end up at large concentrations in the lysosomes [12] and to a lesser extent in the nucleus, particularly in the nucleoli [13]. Therefore, the question naturally arises as to whether the CNDs modify the physiology or metabolism of the lysosomes. Furthermore, when considering CND-mediated drug delivery, one could envisage protocols where the lysosome is the primary therapeutic target, such as enzyme replacement therapy, which has shown successes in the treatment of the quite severe lysosomal storage disorders, as being particularly promising [14].

In this context, it is relevant to clarify how CNDs as well as conjugates formed by CNDs and projectile molecules influence the lysosomal metabolism and trafficking, which comprises such different tasks as degradation of proteins and extracellular particles, nutrient sensing or catabolite export.[15] Furthermore, CNDs may be suitable as carriers for selective drug delivery into lysosomes, while possible carrier effects remain to be evaluated.

In the present study, we expose MCF-7 cells to pristine CNDs as well as to CNDs conjugated to branched polyethylenimine (bPEI) with a molecular weight of 600 Dalton (bPEI-CNDs) up to high concentrations, where the bPEI, a cationic polymer with abundant amine groups, plays the role of a test molecule to be delivered into the lysosome [16, 17]. The bPEI molecule was chosen because it is well established as a vehicle for non-viral gene or drug delivery[18], for example to deliver doxorubicin to the nucleus [19] or to stabilize fragile proteins during transduction [20]. The transfection via bPEI is based on its ability to buffer the pH by binding H^+ ions, which can ultimately lead to rupture of the acidified endosomes or lysosomes, allowing the encapsulated particles to enter the cytosol [16, 21]. Furthermore, bPEI can be regarded as a drug by itself with functions like gastric emptying [22], blockage of fibrin

formation [23] or the enhancement of the permeability of Gram negative bacterial membranes [24]. These wide-spread applications enable conclusions regarding greatly varying aspects of the studied effects.

We characterize the functionality of the lysosomes via monitoring the expression levels or activity of lysosomal markers, such as the enzymes cathepsin B and L, or the autophagy-related markers SQSTM1/p62 and LC3B-II (LC3) in response to the exposures. The multifunctional protein p62 mediates the recruitment of damaged or foreign proteins to the autophagosomal pathway which leads to the lysosomal degradation of the proteins and the autophagy receptor p62 itself [25–27]. Since p62 is degraded during this process, a decrease or increase in its level indicates enhanced or hampered autophagy, respectively. LC3 is another well-established marker for autophagy-related processes, since it is involved in autophagosome formation as well as the binding of p62-marked cargo [28, 29]. Due to its partial degradation during the digestion [30], LC3 levels can be used for sensing the autophagic activity as well. Cathepsines belong to the group of lysosomal hydrolases and are part of a variety of processes including the degradation of proteins [31–33]. By monitoring their activity and the expression levels of the two autophagy markers, conclusions about the status of the lysosomes are drawn, and the effects of the bPEI bound to CNDs on the lysosomal function in comparison to both pristine CNDs as well as free bPEI are specified.

MATERIALS AND METHODS

Preparation of CNDs and bPEI-CND compounds

The CNDs were synthesized according to a modified version of the protocol of Qu et al. [3] which has been described elsewhere [34]. Briefly, 210 mg anhydrous citric acid (Thermo Fisher, #036664.22) and 340 mg Diethylenetriamine (DETA, Merck, #8032740100) are heated in a sealed, teflon-lined microwave vessel under continuous stirring at 180 °C for 2:30 min (CEM Discover). The product was dissolved in 10 ml deionized (DI) water, transferred to a dialysis device (Repligen, Float-A-Lyzer, 0.1-0.5kD, #G235061) and dialysed against 2 l DI water for 48 h with three water exchanges. After dialysis, the product was lyophilized to determine the final mass and dissolved with the targeted concentration in the required solvents for further experiments.

The functionalization of the CNDs was implemented with N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC, Merck, #E6383) and N-Hydroxysuccinimide (NHS, Merck, #130672) coupling. The lyophilized CNDs were solved in DI water in a concentration of $20 \, \text{mg ml}^{-1}$. After solving 1 g of EDC in 5 ml DI water, 10 ml of the CND solution were added and stirred for 10 min. 1 g of NHS was added and stirred for another 10 min before 200 mg of bPEI (Merck, #408719) was added. The final solution was stirred for 24 h. Afterwards the solution was transferred to dialysis devices (Repligen, Float-A-Lyzer, 3.5-5kD, G235065) and dialyzed until the peak of the residual coupling reagents in the 1H-NMR spectrum (around 2.65 ppm) was sufficiently low (dialysis for at least 132 h with two water exchanges per 24 h). The final product was lyophilized to determine the dry mass and used to prepare the solutions for the experiments.

Characterization of the CNDs

Characterization of the unfunctionalized CNDs via fluorescence spectroscopy, Raman spectroscopy, X-ray photoelectron spectroscopy (XPS), CHN chemical elemental analysis and high resolution transmission electron microscopy (HRTEM) has been previously described by Fasbender et al. [2] and led to the following results. CHN chemical elemental analysis revealed that the CNDs consist of 40 % C, 19 % N and 8 % H atoms. The remaining fraction could be assigned to oxygen atoms by XPS analysis. Furthermore, XPS revealed that 29 % of the carbon bonds consist of C-C single bonds, while the remaining bonds were equally assigned to C-O and C-N bonds. Via Raman spectroscopy, it was shown that the C-C bonds consist of sp^2- and sp^3-hybridized carbon atoms. Furthermore, COOH/C-OH and C=O/C-O edge functional groups were identified. The mean particle size was measured via HRTEM and determined to be 3.3 nm with a FWHM of 0.6 nm. The fluorescence spectra revealed absorption peaks at 240 nm and 350 nm which can be assigned to $\pi - \pi^*$ transitions of C=C bonds and to $n - \pi^*$ transitions of C=O bonds, respectively [3]. Excitation between 320 nm and 400 nm leads to a fluorescence emission with a maximum around 460 nm. The quantum yield for excitation at 360 nm was found to be 23 %.

In this work, CNDs and bPEI-CNDs were synthesized as described above and their fluorescence and 1H-NMR spectra were used for characterization. The fluorescence spectra of CNDs, bPEI-CNDs and bPEI in DI water, shown

in Fig. 1, were obtained using a Horiba Duetta™ Fluorescence and Absorbance Spectrometer. Emission spectra were measured with excitation at 360 nm wavelength.
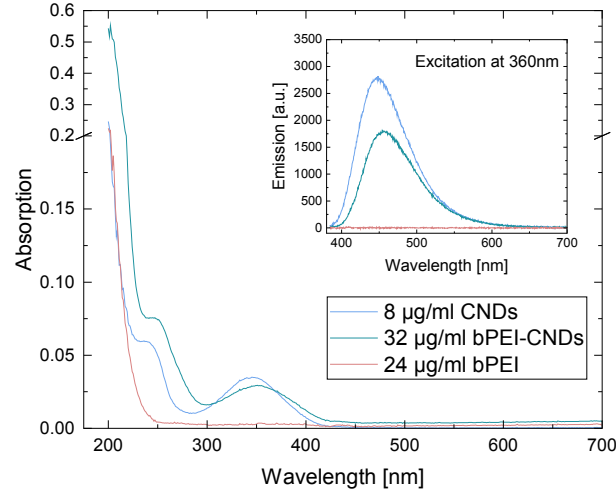

FIG. 1. Main figure: Absorption spectra of the pristine CNDs, bPEI as well as the bPEI-CND conjugate. The inset shows the fluorescence spectra under excitation at 360 nm wavelength. Only the CNDs contribute to the fluorescence detected in this study.

The spectra of the pristine CNDs, shown in Fig. 1, are in good agreement with findings reported elsewhere [3, 34]. They show an absorption peak at approx. 346 nm and a shoulder in the deeper UV-region around 240 nm. After an excitation with 360 nm, the fluorescence signal with a maximum at 449 nm can be observed. Following the functionalization, the fluorescence spectra of the bPEI-CNDs are slightly red shifted, see the inset in Fig. 1. Their absorption peaks appear at 353 nm and 245 nm. The emission maximum is now located at 456 nm. bPEI shows no absorption resonances and does not fluoresce. Furthermore, the functionalization of CNDs leads to a lower absorption at the same CND mass concentration compared to the pristine CNDs. An increase by a factor of 4 for the mass of the bPEI-CND particles results in the same absorption at the maximum around 350 nm as the pristine CNDs, indicating that the CND contributes $\approx 25\%$ of the mass of a bPEI-CND conjugate. Moreover, the fluorescence spectra indicate that a covalent bond between the CNDs and bPEI has been established, most likely via unsaturated sites at the CND edges.

For an independent confirmation of the successful functionalization, ^{1}H-NMR spectra (600 MHz) of CNDs, bPEI-CNDs and bPEI in D_2O were recorded with a Bruker Avance III - 600 by the CeMSA@HHU. The data was then processed and displayed with MestReNova (14.2.0-26256).

As seen in Fig. 2, the ^{1}H-NMR spectrum of pristine CNDs shows narrow compound peaks between 1.99 ppm and 4.85 ppm (detailed figures in the supplement). After the functionalization, the CND peaks are broadened, indicating their successful bond to larger molecules (in this case bPEI). We tentatively explain this by a lower rotational diffusion due to an enlarged particle size. As a consequence, the direct nuclear spin-spin coupling, which is dependent on the angle between the external magnetic field and the binding vector between two atoms, is no longer averaged out, such that the peaks of the bPEI-CNDs appear broadened compared to those of pristine CNDs.

Cell Experiments

Cell Culture

For all experiments, MCF-7 cells were used. The cells were cultured in RPMI 1640 (Biowest, #L0501) with 10 % Fetal Bovine Serum (FBS, Sigma-Aldrich, #F2442), $100\,U\,ml^{-1}$ penicillin and $100\,\mu g\,ml^{-1}$ streptomycin (Sigma-Aldrich, #P0781) and $300\,mg\,l^{-1}$ L-Glutamin (Sigma-Aldrich, #G7513) in an incubator with humidified air at 37 °C and 5 % CO_2. The cells were passaged every three to five days using Trypsin-EDTA solution (Sigma-Aldrich #T3924).
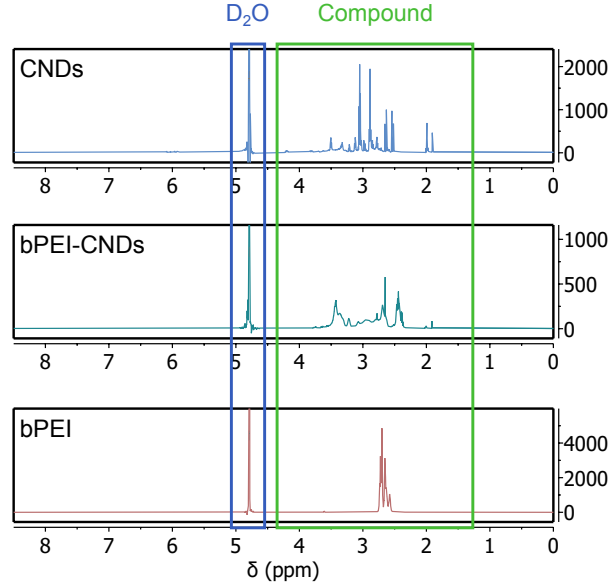
FIG. 2. ^{1}H-NMR spectra of CNDs, bPEI-CNDs and bPEI diluted in D_2O. The D_2O peak of the solvent is visible at 4.79 ppm.

TABLE I. Compound concentrations

Compound	Concentration [mg ml^{-1}]	Justification
CNDs	0.5	no influence on cell viability after 48 h incubation [12]
bPEI-CNDs	0.5	mass-concentration equal to 0.5 mg ml^{-1} CNDs
bPEI-CNDs	2	particle concentration equal to 0.5 mg ml^{-1} CNDs
bPEI	0.1	no influence on cell viability after incubation for 48 h
bPEI	0.333	cell viability is above 75 % after incubation for 48 h as seen in Fig. 4
bPEI	0.375	equal to concentration of bound bPEI in 0.5 mg ml^{-1} bPEI-CND
bPEI	1.5	equal to concentration of bound bPEI in 2 mg ml^{-1} bPEI-CND

Incubation with CNDs, bPEI-CNDs and bPEI

CNDs, bPEI-CNDs and bPEI were dissolved in DPBS (Dulbecco's phosphate-buffered saline, Gibco, #14190144) and sterile filtered through a 0,2 μm PES membrane (Sarstedt, 83.1826.001) prior to their incubation with cells. The compounds were added to the culture medium, such that the final concentrations are obtained. All samples (including the control) were adjusted to contain the same amount of PBS while retaining the correct compound concentration. The cells are then incubated for 48 h at humidified air with 5 % CO_2 and 37 °C. In Table I, the selected compound concentration are listed along with the corresponding justifications.

Confocal Fluorescence Microscopy

For confocal fluorescence microscopy with Lysotracker™Deep Red ("lysotracker", Thermo Fisher, Invitrogen™, #L12492) staining, the cells were seeded in a 8 well μ-Slide (ibidi, ibiTreat #80806) and treated for 48 h as described above. The basal culture medium was exchanged to phenol red-free RPMI 1640 (Biowest, #L0505). After the incubation time of 48 h, the medium was removed and new medium containing 50 nM lysotracker was added. After an incubation for 45 min in the dark at humidified air with 5 % CO_2 and 37 °C the medium was removed and fresh medium without lysotracker was added. The samples were directly imaged using a Zeiss LSM 710 with a Plan-Apochromat 63x/1,4 Oil objective and a closed sample chamber heated to 37 °C. The CNDs were excited with a 405 nm laser

diode and the fluorescence was detected in the range of 410-580 nm ("CND channel"). The lysotracker was excited with a 633 nm HeNe Laser and the emission was detected in the range of 647-754 nm ("lysotracker channel"). The transmitted light images were obtained through the lysotracker channel. All images were obtained using the same measurement parameters. The images were plotted using OMERO.figure (v6.0.1).

Image analysis To determine the amount and size of lysosomes per cell in each sample, the microscopy images were analyzed using Fiji [35] (with ImageJ v1.54f). First, background noise is removed through a convoluted background subtraction (median, radius:10) by the implemented *BioVoxxel* plugin and an intensity threshold is set. Afterwards, the amount and size of the lysosomes is determined using the *Analyze Particles* function (size: 0.01 μm²-Infinity, circularity: 0.00-1.00) provided within ImageJ. The macro used is provided in the supplement.

Viability Assay

The cell viability was measured using an MTT assay (Roth, #4022). The cells were seeded in a 96-well plate in triplicates and incubated as described above. For the positive control, cells were treated with 5 μM staurosporine (Sts, Biozol, #LCL-S-9300) for 48 h. Afterwards, 20 μl MTT solution (5 mg ml^{-1} in DI water) were added to each well and incubated for 40 min. The medium containing residual MTT was removed and 100 μl DMSO were added to each well. The plates were shaken gently until violet formazan was dissolved. The absorbance was measured at 570 nm and 650 nm (for reference) with a microplate reader (SynergyMx, BioTek, Winooski, VT, USA). The reference absorption at 650 nm was subtracted from the absorption at 570 nm. Afterwards the offset of wells without cells was was set to zero. The mean value of the control cells was set to a viability of 100 %.

Immunoblotting/Western Blot

After the incubation with the compounds described above, the cells were washed with DPBS and treated with full culture medium or starvation medium (Earle's Balanced Salt Solution, EBSS, Gibco, #24010-043) containing bafilomycin A_1 (BafA_1, Sigma-Aldrich, #B1793) or a solvent control for 6 h. After the incubation time, the cells were harvested in chilled DPBS by scraping, pelletized at 300 g and 4 °C for 5 min and frozen in liquid nitrogen. After cell lysis in lysis buffer (20 mM Tris–HCl, 150 mM NaCl, 500 μM EDTA, 1 % [v/v] Triton X-100, 1X protease inhibitor cocktail [Roche, #4693132001] and 1XPhosSTOP [Roche, #04906837001]) for 30 min on ice, the lysates were cleared by centrifugation at 18,000 rcf and 4 °C for 15 min and the protein concentration was determined by Bradford assay. Sample buffer was added (62.5 mM Tris, 8.6 % [v/v] glycerol, 2 % [w/v] SDS, 33.3 μg ml^{-1} bromophenol blue, 1 % [v/v] β-mercaptoethanol) and the samples were heated at 95 °C for 5 min. Equal amounts of protein (25 μg) were subjected to SDS-polyacrylamide gels and afterwards transferred to PVDF membranes (Merck, #IPFL00010). The membranes were blocked with 5 % milk powder in TBST and incubated in the indicated primary antibodies (anti-LC3B, CST #2775; p62, PROGEN #gp62-c; Cathespin B, CST #31718; GAPDH, Abcam #ab8245; β-Actin, Sigma-Aldrich #A5316) followed by appropriate IRDye 800- or IRDye 680-conjugated secondary antibodies (LI-COR Biosciences). The fluorescence signals were detected using an Odyssey Infrared Imaging system (LI-COR Biosciences, Lincoln, NE, USA) and signals were quantified using Image Studio (LI-COR Biosciences, Lincoln, NE, USA). The density of each protein band was divided by the average of the density of all bands of this protein on the membrane. The ratios were normalized to the loading control, and fold changes were calculated by dividing each normalized ratio by the average of the ratios of the control line (full medium). For the quantification of Cathepsin B protein levels in Figure S9, both detected protein bands were included in the quantification.

Cathepsin Assay

The Cathepsin B and Cathepsin L activity was measured using the Cathepsin Activity Assay Kits (abcam, #ab65300 and #ab65306) according to the manufacturer's instructions. 10 μg of protein per sample were used in both assays. For inhibitor control, 100 μM Z-FF-FMK (Merck, #219421) was used. The fluorescence of the samples was measured in duplicates with a microplate reader (SynergyMx, BioTek, Winooski, VT, USA) at 505 nm after 405 nm excitation. The average fluorescence intensity of "buffer only" wells without cell lysate was subtracted from the sample fluorescence.

Mass spectrometry (MS)-based proteomics and secretomics

Sample Preparation: After incubation with 0.5 mg ml^{-1} CNDs for 48 h (as described above), the cells were carefully washed 6 times with DPBS. Afterwards, the cells were incubated for 6 h in full cell culture medium, basal medium RPMI1640 or in starvation medium EBSS. Only culturing in basal medium or EBSS was selected for secretome analysis, as additives such as FBS would severely impair the detection of secreted proteins. After the incubation, the conditioned medium (basal medium or EBSS) was collected, centrifuged (5 min, 800 g, 4°C) and filtered through a 0.2 µm membrane (Acrodisc, 32 mm Syringe Filter with 0.2 µm Supor Membrane; Pall, #4652). Aliquots were shock frozen in liquid nitrogen and stored at -80 °C. The cells were washed 3 times with chilled DPBS, harvested in DPBS on ice via scraping and pelleted by centrifugation (5 min, 800 g, 4°C). The supernatant was discarded and the cells were stored at -80°C until further processing. A total of 5 replicates for each condition were prepared.

Proteomics: Proteins were extracted from frozen cell pellets as described elsewhere [36, 37]. Briefly, cells were lysed and homogenized in chaotropic lysisbuffer (30 mM Tris, 2 M thiourea, 7 M urea, 4 % CHAPS, pH 8.5; 5 µl per mg cell wet weight) using a TissueLyser (1 min, 40 Hz; Qiagen) and ultrasound (6x 10 s under ice cooling in an ultrasonic bath). After centrifugation (15 min, 16000 rcf, 4 °C), supernatants were collected. After determination of protein concentration (Pierce 660 nm Protein Assay, Thermo Fischer Scientific, #22662), samples were adjusted to 0.5 mg ml^{-1} total protein concentration with SDS buffer (final 7.5 % glycerol, 3 % SDS, 37.5 mM Tris/HCl pH 7.0). A quality control was performed by SDS-PAGE using 2 µg total protein per condition and replicate, respectively, and silver staining according to Heukeshoven and Dernick [38] with slight modifications. 5 µg total protein per condition and replicate were reduced (final 20 mM dithiothreitol, 20 min, 56 °C), alkylated (4x molar excess iodoacetamide to dithiothreitol, 15 min, r.t., protected from light), quenched (same amount dithiothreitol as for reduction, 15 min, r.t.) and finally underwent tryptic digestion (200 ng trypsin in 20 µl 50 mM triethylammonium bicarbonate) after applying a slightly modified sp3 protocol [39] using 50 µg 1:1 mix Sera-Mag SpeedBeads (GE #45152105050250 and #65152105050250) and final 50 % ethanol for protein precipitation and 80 % ethanol (3x) as well as acetonitrile (1x) for washing the protein-solid phase aggregates. Peptides (25 % of input) were reconstituted in 0.1 % trifluoracetic acid and subjected to LC-MS analysis.

Secretomics: Secretomics was performed as described by Vogt et al. [40]. An aliquot (400 µl) per condition and replicate was thawed on ice in the presence of a protease inhibitor cocktail (added 50 µl of a solution of 1 complete ULTRA tablet, mini, EDTA-free in 2 mL water; Roche, #05892791001), supplemented with SDS buffer (added 50 µl of 30 % glycerin, 12 % SDS, 150 mM Tris/HCl pH 7.0), reduced (added 40.5 µl of 100 mM dithiothreitol; 20 min at 56 °C under shaking), alkylated (added 54 µl of 300 mM iodacetamide; 15 min at r.t. protected from light), and quenched (added 40.5 µl of 100 mM dithiothreitol; 15 min at r.t.). Applying a slightly modified sp3 protocol [39], proteins were precipitated (added 10 µl of 20 mg ml^{-1} 1:1 bead-mix of pre-washed Sera-Mag SpeedBeads [GE #45152105050250 and #65152105050250] in water; added 645 µl ethanol abs. p.a.; 15 min at 24 °C under shaking), washed (3x 80 % ethanol, 1x acetonitrile) and digested (100 ng trypsin in 20 µl 50 mM triethylammonium bicarbonate). Peptides were reconstituted in 0.1 % trifluoracetic acid and subjected to LC-MS analysis.

LC-MS analysis: For the LC-MS analysis, an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher), operated in positive mode and coupled with a nano electrospray ionization source connected with an Ultimate 3000 Rapid Separation liquid chromatography system (Dionex / Thermo Fisher) equipped with an Acclaim PepMap 100 C18 column (75 µm inner diameter, 25 cm length, 2 mm particle size from Thermo Fisher) was applied using a 120 min LC gradient. Capillary temperature was set to 275 °C and source voltage to 1.5 kV. MS survey scans had a mass range from 200 to 2000 m/z at a resolution of 120,000. The normalized AGC target was set to 62.5 % and the maximum fill time was 60 ms. A cycle time of 2 s was employed for isolation and fragmentation of the most intensive peptide ions per survey scan by high-energy collision dissociation (HCD).

Data analysis: MaxQuant (version 2.1.3.0, Max Planck Institute for Biochemistry, Planegg, Germany) was used for peptide and protein identification and quantification using a human sequence database (UniProtKB, downloaded on 01/27/2021, 75777 entries). Methionine oxidation and N-terminal acetylation were considered as variable modifications and carbamidomethylation at cysteine residues as fixed modification. The identification threshold was set as a false discovery rate of 1 % on protein and peptide level. A total of 3239 protein groups were identified after removing potential contaminants, reverse hits, and proteins only identified by modified peptides (only identified by site).

Both, intensities and label free quantification (LFQ) intensities as measures for relative protein abundance by the MaxQuant software (output proteinGroups.txt file) were statistically analyzed using *R* (version 4.2.0). First, the data were normalized such that the median of the logarithmic (LFQ) intensity differences over all proteins between two samples, respectively, approached zero. As a quality control, a principle component analysis (PCA) using the *prcomp()*-function and cluster analyses using the *heatmap()*-function and cluster methods "ward.D", "ward.D2",

"single", "complete", "average", "mcquitty", "median", and "centroid" were performed. Replicate 1 of the proteome control (no CND treatment) sample using EBSS medium was identified as outlier and omitted in further analyses.

Testing for significant differential protein expression (up- or downregulation; proteomes) or differential protein secretion (secretomes) of the CND treated samples vs. the control samples was performed using the "Significance Analysis of Microarrays" (SAM) analysis method [41] within the *Siggenes R* package, separately for each of the three (two for secretomes) media. For this approach, a minimum of four valid values had to be present in at least one group (CND treated or control), data were log 2 transformed to reach a normal distribution like data structure, and missing values were filled in with random values from samplewise downshifted normal distributions (0.3 SD width, 1.8 SD downshift). A permutation based false discovery rate (FDR) of 5 % was used as significance cutoff. Lysosomal lumenal and lysosomal membrane protein classification was performed according to the list of lysosomal proteins contained in Table S3 of the supplementary material of Richards et al. [42].

Data availability: The mass spectrometry proteomics and secretomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [43] partner repository with the dataset identifier PXD053105.

Statistical analysis

The exact number of replicates measured per experiment is indicated in the corresponding figure legends. Viability assay, cathepsin B/L assay and immunoblot data are represented as mean ± SD. The statistical analysis was performed with OriginPro (Version 2021b, OriginLab Corporation). The statistical tests used are stated in the figure legends. In summary, for the viability assay (Fig. 4), cathepsin B/L (Fig. 7) and quantification of the lysotracker (Fig. 8), one-way ANOVA with Bonferroni comparisons was used. For the immunoblots of CNDs and bPEI-CNDs (Fig. 6), two-way ANOVA with Bonferroni comparisons was used. For the immunoblots of bPEI (see supplement), three-way ANOVA with Bonferroni comparisons was used. For MS-based proteomics analysis, SAM and a FDR < 5 % was used as described above.

RESULTS AND DISCUSSION

CNDs and bPEI-CNDs accumulate within lysosomes

We begin by characterizing the distribution of the CNDs and of the bPEI-CNDs within the cells. In Fig. 3, microscopy images of the MCF-7 cells after CND - as well as after bPEI-CND - incubation are shown. For lysosomal staining, lysotracker was used. By merging the cyan (bPEI-) CND channel and the magenta lysotracker channel (Fig. 3 (g) and (k)), it emerges from the appearance of the mixed color that (bPEI-) CNDs accumulate mainly in the lysosomes, in accordance with earlier reports [12, 34]. Due to imaging of living cells with moving organelles and a time delay between the measurement of the two fluorescence channels, a slight displacement of the overlaying CND and lysotracker signals may be observed.

We emphasize that a weak signal of the pristine CNDs is visible across the cell, possibly indicating a small rate of CND leakage from the endolysosomal pathway. A background signal from bPEI-CND signal, on the other hand, in other cell compartments besides lysosomes is not detectable. Apparently, the bPEI molecules localize the CNDs inside the lysosomes. We tentatively explain this by the enlarged particle size and / or by an altered charge of the bPEI-CNDs preventing lysosomal escape. Alternatively, the bPEI may buffer or change the lysosomal pH [44] and thus its functionality. Nevertheless, the microscopy images confirm a successful transport of functionalized CNDs, i.e. the transport of the compound bPEI-CND into the lysosomes and therefore support the potential application of CNDs as carriers for lysosomal-targeting therapeutics.

In contrast to free bPEI, CNDs and bPEI-CNDs do not affect the cell viability

Since pristine bPEI does not fluoresce, only indirect conclusions about its intracellular distribution can be drawn from functional studies. The first of these studies comprises viability assays. Previous experiments have shown that an incubation with CNDs in a concentration of 0.5 mg ml^{-1} for 48 h does not alter the viability [12]. We confirm this
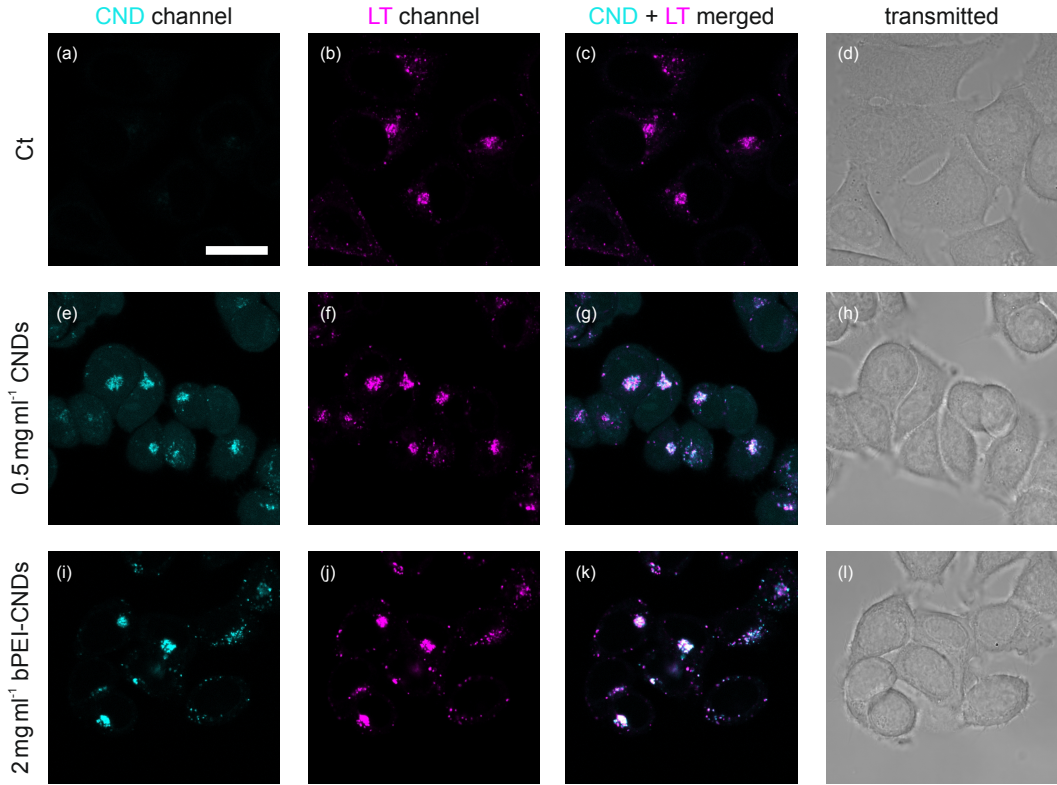
FIG. 3. Confocal microscopy images of MCF-7 cells incubated with CNDs or bPEI-CNDs. Lysosomes are stained with lysotracker. The CND channel is shown in cyan, the lysotracker ("LT") channel in magenta. (a-d) MCF-7 cells without (bPEI-) CND incubation as a control. (e-h) MCF-7 cells incubated with $0.5\,\mathrm{mg\,ml^{-1}}$ CNDs. (i-l) MCF-7 cells incubated with $2\,\mathrm{mg\,ml^{-1}}$ bPEI-CNDs. Merging of the cyan CND channel and magenta LT channel results in blue-white color-signals. As seen in the CND-LT-overlay images (g) and (k), CNDs and bPEI-CNDs mainly accumulate inside the lysosomes. The scale bar of $20\,\mathrm{\mu m}$ shown in (a) applies to all images.

observation in our system and complement it by investigating the influence on the cell viability of the functionalized CNDs as well as the influence of unbound bPEI. As can be seen in Fig. 4 (a), a concentration up to $2\,\mathrm{mg\,ml^{-1}}$ of bPEI-CNDs does not impact the viability significantly. Concentrations of free bPEI above $0.333\,\mathrm{mg\,ml^{-1}}$, however, do reduce the cell viability significantly, see Fig. 4 (b). A free bPEI concentration of $0.375\,\mathrm{mg\,ml^{-1}}$ equals the concentration of the bound bPEI in the $0.5\,\mathrm{mg\,ml^{-1}}$ bPEI-CNDs solution and causes already a viability decrease to $68\,\%$. In addition, it is observed that a free bPEI concentration of $1.5\,\mathrm{mg\,ml^{-1}}$ which corresponds to a bound bPEI concentration in $2\,\mathrm{mg\,ml^{-1}}$ bPEI-CNDs leads to a viability of $14\,\%$ which is similar to that one observed after incubation with the apoptosis-inducing and positive control staurosporine.

We identify two possible reasons for this weakening effect of the CNDs on the bPEI-induced effects. First, the binding of bPEI to CNDs will lead to capture of bPEI inside the lysosomes and thus to a reduction of cytotoxic effects of bPEI outside of the lysosomes. This is plausible considering the reported effects bPEI showed in other studies when inducing the lysosomal pathway [45]. This explanation is in good accordance to the microscopy images (Fig. 3) which show a strong localization of bPEI-CNDs inside the endo-lysosomal pathway. Second, possible cytotoxic effects by the bPEI inside the lysosomes may be attenuated due to saturation of otherwise reactive bPEI sites by their bonds to the CNDs. We will revisit this explanation below.
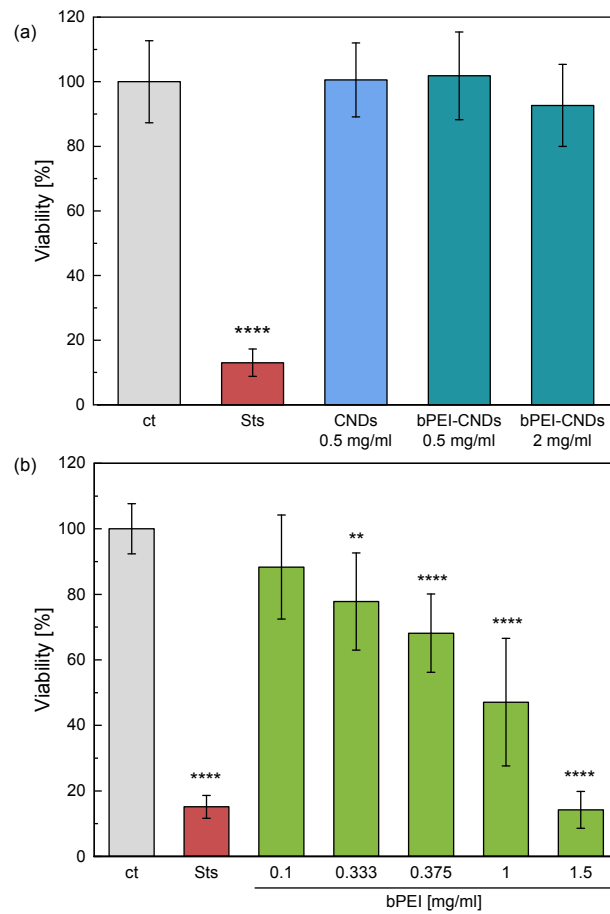
FIG. 4. Viability of MCF-7 cells after incubation with (a) CNDs and bPEI-CNDs or (b) bPEI. Staurosporine (Sts) was used as a positive control. Data are represented as mean $\pm$ SD of N=3 biologicalreplicates with each n=3 technicalreplicates (expect bPEI $1\,\mathrm{mg\,ml^{-1}}$: N=2, n=3). p values were determined by one-way ANOVA with Bonferroni comparison. In comparison with the control sample "ct": **p<0.01, ****p<0.0001.

CNDs do not influence the cells' proteome

To investigate the influence of the pristine CNDs on the cells' proteome, MS-based quantitative differential (CND incubation vs. control) proteomics and secretomics analyses were performed using three (two for secretomics) different media (complete cell culture medium, basal medium RPMI1640 or starvation medium EBSS), resulting in a total of 3239 identified and quantified proteins. For the statistical analyses, the two values for relative protein quantification given by the MaxQuant software, the "intensities" and the "LFQ intensities", were used (The "intensities" simply comprise the sum of all intensities of the identified peptides of a given protein in a given sample, while the "LFQ intensities" are calculated based on the intensities of the peptides that share their identification in the samples to be compared. Therefore, LFQ intensity values tend to be more accurate when there are enough shared peptides, but disregard many values from non-shared peptides, resulting in more missing data [46]).

Except for membrane-associated progesterone receptor component 2 (*PGRMC2*) in the proteome for EBSS medium, no protein showed consistent significance (permuation-based FDR < 5 %, SAM analysis) for both the intensity- and LFQ intensity-based statistical analyses in neither full medium nor basal RPMI1640 medium, and there was no protein that was significant over all analyses (see Fig. 5 for an exemplary volcano plot and Fig. S6 for the full set). Notably, none of the red or blue marked proteins associated with the lysosomal lumen and lysosomal membrane, respectively, were found to be significantly altered. Moreover, no profound functional connections or corresponding pathways were found for the significant proteins for the different conditions by a functional enrichment analysis using the STRING
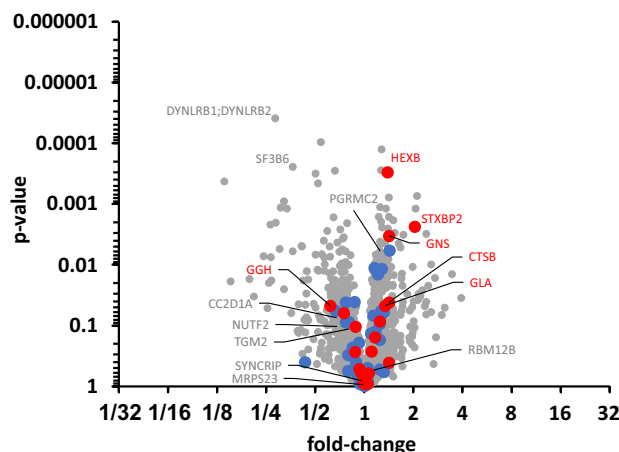
FIG. 5. Volcano plot based on LFQ intensity values for MS-based proteomics. Analysis of differential protein abundance (fold-change) after CND treatment of MCF-7 cells in full culture media was performed. See Fig. S6 for the full set of proteome and secretome analyses in the three (two for secretomes) tested media. No proteins with a permutation-based FDR (SAM analysis) $< 5\%$ were identified in the present analysis. Grey gene name labels indicate proteins, for which this criterion was met for other conditions. Lysosomal lumenal and lysosomal membrane proteins are labelled with blue and red data points, respectively, and lumenal proteins with p-value < 0.05 are labelled with red gene names.

database [47] (see Fig. S7).

In the secretome data, no significantly altered proteins were detected (permuation-based FDR $< 5\%$, SAM analysis), indicating that pristine CNDs do not strongly alter the cellular secretion of proteins, and therefore also suggesting that pristine CNDs are inert to the metabolic processes of the cells.

CNDs do not influence lysosomal functions, and bPEI-CNDs show attenuated effects compared to free bPEI

We proceed by functional studies of the lysosomes after exposure to the nanoparticles, exemplified by their autophagic activity via the markers p62 and LC3, as well as the activity of the cathepsins B and L, as a measure of the rate of lysosome-based degradation of proteins. Here, we keep the nanoparticle concentrations in the regime where their effect on the viability is below the significance threshold or the viability of the cells is still above 75%.

In the full cell culture medium ("RPMI") we do not observe any significant effect of CNDs on p62 and LC3 levels, see the left hand side in Fig. 6 (a). At high bPEI-CND concentrations ($2\,\mathrm{mg\,ml^{-1}}$) the bPEI-CNDs lead to a significant decrease of the p62 level to 31% of the control, and a tendency towards decreasing p62 concentrations is already noticeable for the concentration of ($0.5\,\mathrm{mg\,ml^{-1}}$). The LC3 levels are not significantly altered, nevertheless a tendency to increased LC3 levels can be seen in Fig. 6 (b) for bPEI-CND treatments. Starvation (right hand side, "EBSS") reduces p62 levels, and this effect is not significantly altered by any additional treatment. Again, increased LC3 levels can be observed for bPEI-CND treatments. Collectively, CNDs do not appear to affect cellular autophagic capacity. In order to further characterize the observed effects for bPEI-CNDs, we combined bPEI (unbound) with bafilomycin A_1. This compound is an inhibitor of the vacuolar H^+-ATPase and ultimately inhibits the fusion of autophagosomes with lysosomes [48]. The combination of a stimulus $+/-$ bafilomycin A_1 allows the analysis of the autophagic flux. As shown in Fig. S8, bafilomycin A_1 cannot prevent the bPEI-induced reduction of p62, indicating that autophagy-unrelated mechanisms contribute to this effect. Since the effects of unbound bPEI and bPEI-CNDs on p62 levels are similar, we assume that these autophagy-unrelated mechanisms also cause the effects on p62 and LC3 levels shown in Fig. 6. Note that for unbound bPEI a reduced concentration by a factor of 4.5 compared to bound bPEI is needed to effect the cells metabolism significantly.

The activity of cathepsins B and L is not significantly altered upon CND incubation as seen in Fig. 7, which supports the suitability of CNDs as inert carriers for drug delivery systems. After incubation with high bPEI-CND concentrations ($2\,\mathrm{mg\,ml^{-1}}$), cellular cathepsin B activity is significantly enhanced by a factor of 3.3. Incubation with ($0.333\,\mathrm{mg\,ml^{-1}}$) bPEI leads to a significantly increased cathepsin B activity by a factor of 5.8. Therefore bPEI incubation results in a higher cathepsin B activity regardless of the state (bound/unbound). Note that the effect of unbound
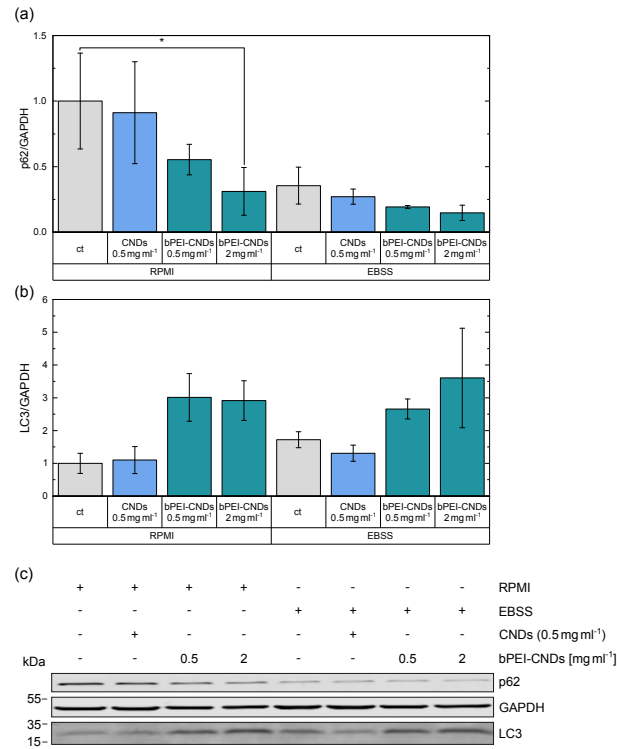
FIG. 6. Influence of CNDs and bPEI-CNDs on cellular (a) p62 and (b) LC3-II levels. Samples were measured in biological triplicates. The levels were quantified and normalized to GAPDH. Data are represented as mean $\pm$ SD. A representativ immunoblot is shown in (c). p values were determined by two-way ANOVA with Bonferroni comparison. Shown is the significance compared to the control-sample ("ct") in the same medium. *$p < 0.05$.

bPEI on the cells cathepsin B activity is attenuated upon CND binding. A concentration of $0.333 \, \mathrm{mg \, ml^{-1}}$ unbound bPEI leads to a higher cathepsin B activity than a concentration of $2 \, \mathrm{mg \, ml^{-1}}$ bPEI-CNDs which corresponds to approximate $1.5 \, \mathrm{mg \, ml^{-1}}$ bound bPEI. The increased cathepsin B activities due to bPEI are in good accordance with an significantly increased level of cathepsin B inside the cells after free-bPEI incubation as shown in the immunoblots in Fig. S9. This could be due to a general higher activity inside the lysosomes, as well as a larger total quantity due to a higher lysosomal number needed due to bPEI-mediated effects on the whole cell. Compared to the cathepsin B activity, for the cathepsin L activity, a concentration dependent effect of bPEI is observed. Lower concentrations of bPEI (bound in $0.5 \, \mathrm{mg \, ml^{-1}}$ bPEI-CNDs or unbound $0.1 \, \mathrm{mg \, ml^{-1}}$) lead to a significant decrease of the relative activity to 58 % respective 44 %. Upon bPEI concentration increase (e.g. bound in $2 \, \mathrm{mg \, ml^{-1}}$ bPEI-CNDs or unbound $0.333 \, \mathrm{mg \, ml^{-1}}$), the cathepsin L activity is significantly enhanced to 125 % or back at an unaltered level. We assume that two different effects occur depending on the bPEI concentration, especially on the free reaction sites of bPEI. At low bPEI concentrations, a potential change of the pH value inside the lysosomes towards neutral could be a reason for the decreased cathepsin L activity as it was seen that bPEI can prevent the strong acidification of lysosomes [44]. Due to the cathepsin L activity optimum in acidic environment, at higher pH their activity is expected to be reduced. This is in good accordance with the unaltered cathepsin B activity at low bPEI concentrations since cathepsins B are known to show still good activity at less acidic environments [49]. Note that the change of the lysosomal pH upon bPEI incubation is still under debate [50], such that other lysosome-affecting effects like direct interactions with unsaturated bPEI-sites need to be considered as well. Since the cathepsin L activity is increased again to a normal or higher level at higher bPEI concentrations, a second bPEI-mediated effect should be considered. Taking the increased cathepsin B and L activity together, we suggest that the increased bPEI levels lead to cellular stress reactions, resulting in higher lysosomal activity. This is in good accordance with the immunoblotting and viability results which indicate autophagy-unrelated mechanisms as described above.
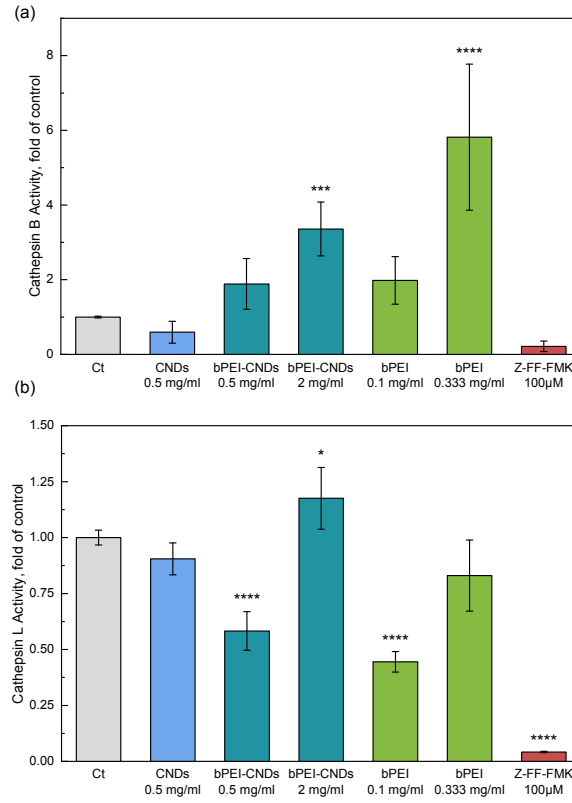
FIG. 7. Activity of (a) cathepsin B and (b) cathepsin L after incubation with CNDs, bPEI-CNDs and bPEI. The cathepsin inhibitor Z-FF-FMK was used as a positive control. The samples were measured in biological triplicates with technical dublicates each. Data are represented as mean ± SD. p values were determined by one-way ANOVA with Bonferroni comparison. In comparison with the control sample "ct": *$p<0.05$, ***$p<0.001$, ****$p<0.0001$.

Lysosomal size and abundance does not change upon CND incubation but is influenced by free bPEI and bPEI-CNDs

Finally, we look at the size distributions of the lysosomes after incubation, see Fig. 8. In agreement with all other observations, there is again no effect of the pristine CNDs on the number (Fig. 8 (a)) and the size (Fig. 8 (b)) of the lysosomes. Incubation of bPEI-CNDs leads to significantly less lysosomes per cell with larger average sizes. For pristine bPEI, concentration dependent effects are observed. Free bPEI causes the lysosomes' size to increase and number of lysosomes to decrease significantly at intermediate concentrations, similar to the effects of bPEI-CNDs. For large bPEI concentrations, the lysosomal size decreases again to a not significantly altered level compared to the control while the average number of lysosomes increases significantly compared to the control. Note that much smaller concentrations are needed for free bPEI (here $0.1 \, \mathrm{mg \, ml^{-1}}$) to qualitatively reach the same effect as bPEI-CNDs. The effect of bPEI is, as seen in the other experiments as well, attenuated due to the binding to CNDs. An increased lysosomal size upon free bPEI or bPEI-CND incubation might be explained by an early-state osmotic swelling as a consequence of the proton buffering capacitance ("proton sponge effect") of bPEI [51]. Rupture of Lysosomes, usually observed when bPEI is used for transfection [52], is unlikely due to missing signal distribution of bPEI-CNDs across the whole cell as seen in the microscopy images (Fig. 3). If other effects like an enhanced lysophagy because of bPEI-damaged lysosomes occur needs to be further investigated. A different lysosomal positioning in the cell could
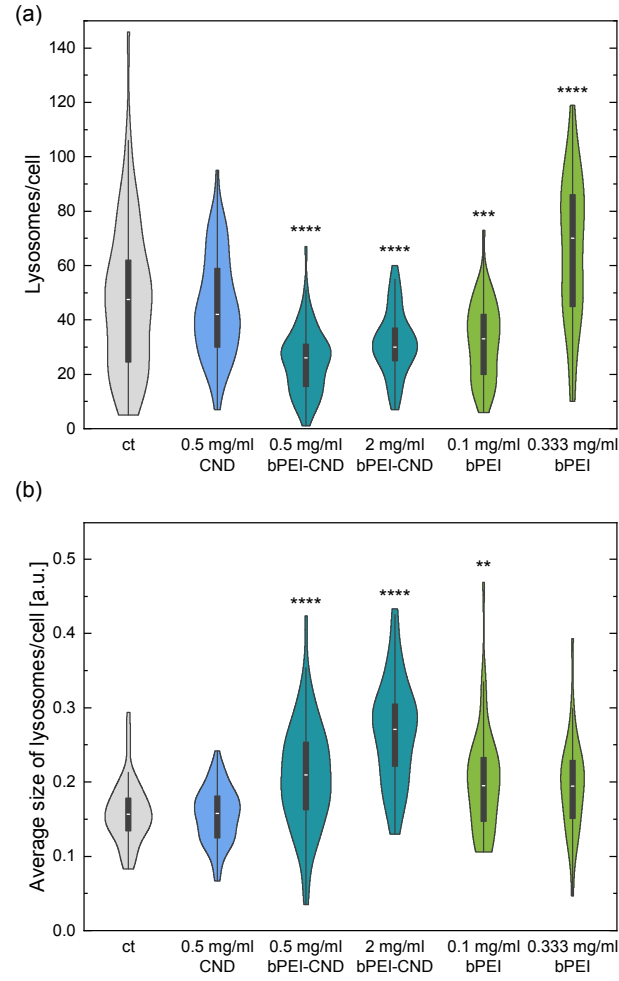
FIG. 8. (a) Number of lysosomes per cell and (b) average size of lysosomes per cells as quantified from the microscopy data (e.g. Fig. 3 and Fig. S4). 34 to 72 cells per condition were analyzed. p values were determined by one-way ANOVA with Bonferroni comparison. In comparison with the control sample "ct": **p<0.01, ***p<0.001, ****p<0.0001.

also be the reasons for the altered values. Since the lysosomal sizes are in the scale of the microscopic resolution, for lysosomes that cluster perinuclear, as seen upon lysosomal perturbations [53], the microscopic distinction could be difficult. This can lead to distorted values since multiple smaller lysosomes could be detected as one. For higher free bPEI concentration, the increased number of lysosomes and unaltered size compared to the control, could be assigned to a general stress reaction of the cell against bPEI. This is also consistent with the previously discussed elevated cathepsin B activity and abundance, as well as slightly increased LC3 levels upon incubation with higher free bPEI concentrations, all suggesting the formation of new lysosomes.

Conclusion

We investigated the effects of pristine fluorescent carbon nanodots on the cells' proteome and on general aspects of the lysosomal activity. We found that even though high numbers of CNDs accumulate inside the lysosomes, they did not affect the lysosomal function nor the overall proteome. We furthermore examined the functionalizability of those CNDs by binding a relevant polymer (bPEI) to them and testing whether the CNDs can transport those compounds into the lysosomes as a target. We have seen indeed that the functionalized CNDs show a more correlated uptake into

the lysosomes compared to the pristine CNDs. The results suggest that the binding to CNDs localizes the bPEI more targeted into lysosomes compared to the distribution of free bPEI inside the cells, resulting in better cell viability and probably less unwanted damage to other cell compartments. Moreover we saw that the bound bPEI could still interfere the lysosomal metabolism as investigated by the autophagy markers LC3 and p62, the cathepsin B and L activity as well as seen by the influenced lysosomal size and quantity. We conclude that the CNDs are suitable as inert carriers in drug delivery systems and can transport the compound to be delivered precisely into the (endo-)lysosomal compartments without interfering with the overall viability of the cells. This could be particularly relevant for the treatment of diseases where the cell should remain intact, such as lysosomal storage diseases or neuro-degenerative diseases, but lysosomal functions should be altered. In such protocols, a possible attenuation of the therapeutic effects due to the binding of the compound to the CND needs to be assessed.

Limitations of the study

All studies reported above have been observed on one cell line. However, since in earlier studies, the uptake of CNDs by live cells has shown only marginal cell-type specific variations, we expect that the behavior will be similar in other cell lines. It remains to be seen in future studies to what extent these lysosomal effects depend on the type/nature of the functional molecule attached to the CND.

Acknowledgements

Author contributions

C.S., C.D., L.B., C.N., T.H. and B.S. conceived the experiments and analyzed and discussed the data. C.S. prepared and characterized the carbon nanodots and the conjugates, prepared the samples for the cell experiments, carried out the assays, some confocal microscopy measurements, the image analysis as well as the statistical analyses and figure preparation. C.D. and L.B. performed the immunoblot analyses and the cathepsin assays. C.N. assisted in the carbon nanodot functionalization and sample preparation for cell experiments as well as performed some confocal microscopy measurements. T.L. and K.S. developed the of MS methodology. K.S. performed funding, supervision and provision of MS infrastructure. T.L. performed MS investigation and sample workup, MS data analysis and curation, and interpretation and visualization of MS data. C.S. and T.H. wrote the first draft of the manuscript. C.D., L.B., C.N., T.L. and B.S. contributed to writing, reviewing and editing of the final manuscript. B.S. and T.H. supervised the project.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Composed of Document S1 and Figures S1-S9. It provides additional information regarding the characterization of the CNDs and the bPEI-CND conjugates, the uptake and intracellular distribution of the nanoparticles, the proteome and secretome analysis and the expression levels of p62, LC3 as well as cathepsin B.

* thomas.heinzel@hhu.de

† bjoern.stork@hhu.de

[1] F. Chen, W. Gao, X. Qiu, H. Zhang, L. Liu, P. Liao, W. Fu, and Y. Luo, Graphene quantum dots in biomedical applications: Recent advances and future challenges, Frontiers in Laboratory Medicine **1**, 192 (2017).

[2] S. Fasbender, L. Zimmermann, R.-P. Cadeddu, M. Luysberg, B. Moll, C. Janiak, T. Heinzel, and R. Haas, The Low Toxicity of Graphene Quantum Dots is Reflected by Marginal Gene Expression Changes of Primary Human Hematopoietic Stem Cells, Sci Rep. **9**, 12028 (2019).

[3] D. Qu, M. Zheng, J. Li, Z. Xie, and Z. Sun, Tailoring color emissions from N-doped graphene quantum dots for bioimaging applications, Light Sci Appl **4**, e364 (2015).

[4] Z. L. Wu, M. X. Gao, T. T. Wang, X. Y. Wan, L. L. Zheng, and C. Z. Huang, A general quantitative ph sensor developed with dicyandiamide n-doped high quantum yield graphene quantum dots, Nanoscale **6**, 3868 (2014).

[5] S. Zhu, Y. Song, X. Zhao, J. Shao, J. Zhang, and B. Yang, The photoluminescence mechanism in carbon dots (graphene quantum dots, carbon nanodots, and polymer dots): current state and future perspective, Nano Res. **8**, 355 (2015).

[6] P. Lesani, G. Singh, C. M. Viray, Y. Ramaswamy, D. M. Zhu, P. Kingshott, Z. Lu, and H. Zreiqat, Two-photon dual-emissive carbon dot-based probe: deep-tissue imaging and ultrasensitive sensing of intracellular ferric ions, ACS Appl. Mater. Interfaces **12**, 18395 (2020).

[7] X. Sun, Z. Liu, K. Welsher, J. T. Robinson, A. Goodwin, S. Zaric, and H. Dai, Nano-graphene oxide for cellular imaging and drug delivery, Nano Res. **1**, 203 (2008).

[8] L. He, T. Wang, J. An, X. Li, L. Zhang, L. Li, G. Li, X. Wu, Z. Su, and C. Wang, Carbon nanodots@ zeolitic imidazolate framework-8 nanoparticles for simultaneous ph-responsive drug delivery and fluorescence imaging, CrystEngComm **16**, 3259 (2014).

[9] I. J. Gomez, B. Arnaiz, M. Cacioppo, F. Arcudi, and M. Prato, Nitrogen-doped carbon nanodots for bioimaging and delivery of paclitaxel, J. Mater. Chem. B **6**, 5540 (2018).

[10] A. Cadranel, J. T. Margraf, V. Strauss, T. Clark, and D. M. Guldi, Carbon nanodots for charge-transfer processes, Accounts of Chemical Research **52**, 955 (2019).

[11] R. Knoblauch and C. D. Geddes, Carbon nanodots in photodynamic antimicrobial therapy: A review, Materials **13**, 4004 (2020).

[12] D. Kersting, S. Fasbender, R. Pilch, J. Kurth, A. Franken, M. Ludescher, J. Naskou, A. Hallenberger, C. v. Gall, C. J. Mohr, R. Lukowski, K. Raba, S. Jaschinski, I. Esposito, J. C. Fischer, D. Fehm, D. Niederacher, H. Neubauer, and T. Heinzel, From in vitro to ex vivo: subcellular localization and uptake of graphene quantum dots into solid tumors, Nanotechnology **30**, 395101 (2019).

[13] X. Wang, Y. Wang, H. He, X. Chen, X. Sun, Y. Sun, G. Zhou, H. Xu, and F. Huang, Steering graphene quantum dots in living cells: lighting up the nucleolus, J. Mater. Chem. B **4**, 779 (2016).

[14] S. R. Bonam, F. Wang, and S. Muller, Lysosomes as a thrapeutic target, Nat Rev. Drug Discov. **18**, 923 (2019).

[15] H. Xu and D. Ren, Lysosomal Physiology, Annu Rev Physiol. **77**, 57 (2015).

[16] O. Boussif, F. Lezoualc, M. A. Zanta, M. D. Mergny, D. Sherman, B. Demeneix, and J. A. Behr, A versatile vector for gene and oligonucteotide transfer into cells in culture and in vivo, Proc Nat Acad Sci USA **92**, 7297 (1995).

[17] P. Vicennati, A. Giuliano, G. Ortaggi, and A. Masotti, Polyethylenimine in medical chemistry, Current Med. Chem. **15**, 2826 (2008).

[18] C. Zhao and B. Zhou, Polyethyleneimine - based drug delivery systems for cancer diagnostics, J Funct Biomater. **14**, 12 (2022).

[19] P. Xu, E. A. Van Kirk, Y. Zhan, W. J. Murdoch, M. Radosz, and Y. SHen, Targeted charde-reversal nanoparticles for nuclear drug delivery, Angew. Chem.Int. ed. **46**, 4999 (2007).

[20] M. Kitazoe, H. Murata, J. Futami, T. Maeda, M. Sakaguchi, M. Miyazaki, M. Kosaka, H. Tada, M. Seno, N. Huh, M. Namba, M. Nishikawa, J. Maeda, and H. Yamada, Protein transduction assisted by polyethyleneimine-cationized carrier proteins, J Biochem. **137**, 693 (2005).

[21] D. W. Pack, A. S. Hoffman, S. Pun, and P. S. Stayton, Design and development of polymers for gene delivery, Nat Rev Drug Discov **4**, 581 (2005).

[22] S. Melamed, G. R. Carlson, J. N. Moss, E. J. Belair, and M. F. Tansy, Gi pharmacology of polyethylemeiminei: effects of gastric emptying in rats, J. Pharm. Sci. **66**, 899.

[23] A. J. Chu, S. Beydoun, S. Mathews, and T. Hoang, Novel anticoagulant polyethylenimine: inhibition of thrombin-catalyzed fibrin formation, Arch. Biochem. Biophys. **415**, 101 (2003).

[24] I. Helander, H.-L. Alakomi, K. Latva-Kala, and P. Koski, Polyethyleneimine is an effective permeabilizer of gram-negative bacteria, Microbiology **143**, 3193 (1997).

[25] Y. Ichimura and M. Komatsu, Selective degradation of p62 by autophagy, Semin Immunopathol **32**, 431 (2010).

[26] S. Ma, I. Y. Attarwala, and X.-Q. Xie, Sqstm1/p62: A potential target for neurodegenerative disease, ACS Chem. Neurosci. **10**, 2094 (2019).

[27] J. D. Lane, V. I. Korolchuk, J. T. Murray, T. Lamark, S. Svenning, and T. Johansen, Regulation of selective autophagy: the p62/SQSTM1 paradigm, Essays Biochem **61**, 609 (2017).

[28] M. B. Schaaf, T. G. Keulers, M. A. Vooijs, and K. M. Rouschop, Lc3/gabarap family proteins: autophagy-(un) related functions, Faseb **30**, 3961 (2016).

[29] S. J. Cherra, S. M. Kulich, G. Uechi, M. Balasubramani, J. Mountzouris, B. W. Day, and C. T. Chu, Regulation of the autophagy protein LC3 by phosphorylation, J Cell Biol **190**, 533 (2010).

[30] T. R. G. Amber N. Hale, Dan J. Ledbetter and I. Edmund B. Rucker, Autophagy, Autophagy **9**, 951 (2013).

[31] T. Yadati, T. Houben, A. Bitorina, and R. Shiri-Sverdlov, The ins and outs of cathepsins: Physiological function and role in disease management, Cells **9**, 1679 (2020).

[32] U. Repnik, V. Stoka, V. Turk, and B. Turk, Lysosomes and lysosomal cathepsins in cell death, Biochim Biophys Acta **1824**, 22 (2012).

[33] M. Scarcella, D. d'Angelo, M. Ciampa, S. Tafuri, L. Avallone, L. M. Pavone, and V. De Pasquale, The key role of lysosomal protease cathepsins in viral infections, International Journal of Molecular Sciences **23**, 9089 (2022).

[34] C. Wimmenauer and T. Heinzel, Identification of nanoparticles as vesicular cargo via airy scanning fluorescence microscopy and spatial statistics, Nanoscale Adv. **5**, 3512 (2023).

[35] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, *et al.*, Fiji: an open-source platform for biological-image analysis, Nat methods **9**, 676 (2012).

[36] L. Sinatra, J. Yang, J. Schliehe-Diecks, N. Dienstbier, M. Vogt, P. Gebing, L. M. Bachmann, M. Sönnichsen, T. Lenz, K. Stühler, A. Schöler, A. Borkhardt, S. Bhatia, and F. K. Hansen, Solid-phase synthesis of cereblon-recruiting selective histone deacetylase 6 degraders (hdac6 protacs) with antileukemic activity, J. Med. Chem. **65**, 16860 (2022).

[37] G. Poschmann, K. Seyfarth, D. Besong Agbo, H.-W. Klafki, J. Rozman, W. Wurst, J. Wiltfang, H. E. Meyer, M. Klingenspor, and K. Stühler, High-fat diet induced isoform changes of the parkinson's disease protein dj-1, Journal of Proteome Research **13**, 2339 (2014).

[38] J. Heukeshoven and R. Dernick, Improved silver staining procedure for fast staining in phastsystem development unit. i. staining of sodium dodecyl sulfate gels, Electrophoresis **9**, 28 (1988).

[39] C. S. Hughes, S. Moggridge, T. Müller, P. H. Sorensen, G. B. Morin, and J. Krijgsveld, Single-pot, solid-phase-enhanced sample preparation for proteomics experiments, Nat Protoc **14**, 68 (2019).

[40] M. Vogt, N. Dienstbier, J. Schliehe-Diecks, K. Scharov, J.-W. Tu, P. Gebing, J. Hogenkamp, B.-S. Bilen, S. Furlan, D. Picard, M. Remke, L. Yasin, D. Bickel, M. Kalia, A. Iacoangeli, T. Lenz, K. Stühler, A. A. Pandyra, J. Hauer, U. Fischer, R. Wagener, A. Borkhardt, and S. Bhatia, Co-targeting hsp90 alpha and cdk7 overcomes resistance against hsp90 inhibitors in bcr-abl1+ leukemia cells, Cell Death Dis **14**, 799 (2023).

[41] V. G. Tusher, R. Tibshirani, and G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, Proc Natl Acad Sci USA **98**, 5116 (2001).

[42] C. M. Richards, S. Jabs, W. Qiao, L. D. Varanese, M. Schweizer, P. R. Mosen, N. M. Riley, M. Klüssendorf, J. R. Zengel, R. A. Flynn, A. Rustagi, J. C. Widen, C. E. Peters, Y. S. Ooi, X. Xie, P.-Y. Shi, R. Bartenschlager, A. S. Puschnik, M. Bogyo, C. R. Bertozzi, C. A. Blish, D. Winter, C. M. Nagamine, T. Braulke, and J. E. Carette, The human disease gene lyset is essential for lysosomal enzyme transport and viral infection, Science **378**, eabn5648 (2022).

[43] Y. Perez-Riverol, J. Bai, C. Bandla, D. García-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma, and J. Vizcaíno, The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences, Nucleic Acids Research **50**, D543 (2021).

[44] S. Roy, D. Zhu, W. J. Parak, and N. Feliu, Lysosomal proton buffering of poly(ethylenimine) measured in situ by fluorescent ph-sensor microcapsules, ACS Nano **14**, 8012 (2020).

[45] N. Mohammadi, N. Fayazi Hosseini, H. Nemati, H. Moradi-Sardareh, M. Nabi-Afjadi, and G. A. Kardar, Revisiting of Properties and Modified Polyethylenimine-Based Cancer Gene Delivery Systems, Biochem Genet. **62**, 18 (2024).

[46] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlfq*, Molecular Cellular Proteomics **13**, 2513 (2014).

[47] D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. T. Doncheva, S. Pyysalo, P. Bork, L. J. Jensen, and C. von Mering, The string database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest, Nucleic Acids Research **51**, D638 (2023).

[48] C. Mauvezin and T. P. Neufeld, Bafilomycin a1 disrupts autophagic flux by inhibiting both v-atpase-dependent acidification and ca-p60a/serca-dependent autophagosome-lysosome fusion, Autophagy **11**, 1437 (2015).

[49] M. C. Yoon, A. Solania, Z. Jiang, M. P. Christy, S. Podvin, C. Mosier, C. B. Lietz, G. Ito, W. H. Gerwick, D. W. Wolan, G. Hook, A. J. O'Donoghue, and V. Hook, Selective neutral ph inhibitor of cathepsin b designed based on cleavage preferences at cytosolic and lysosomal ph conditions, ACS Chemical Biology **16**, 1628 (2021).

[50] R. V. Benjaminsen, M. A. Mattebjerg, J. R. Henriksen, S. M. Moghimi, and T. L. Andresen, The possible "proton sponge" effect of polyethylenimine (pei) does not include change in lysosomal ph, Mol Ther. **21**, 149 (2013).

[51] A. El-Sayed, S. Futaki, and H. Harashima, Delivery of macromolecules using arginine-rich cell-penetrating peptides: ways to overcome endosomal entrapment, AAPS J **11**, 13 (2009).

[52] A. P. Pandey and K. K. Sawant, Polyethylenimine: A versatile, multifunctional non-viral vector for nucleic acid delivery, Mater Sci Eng C **68**, 904 (2016).

[53] J. Pu, C. M. Guardia, T. Keren-Kaplan, and J. S. Bonifacino, Mechanisms and functions of lysosome positioning, J Cell Sci **129**, 4329 (2016).

Supplementary Materials

Lysosomal activity in response to incubation of pristine and functionalized carbon nanodots: Supplementary material

Carla Sprengel ®,[1] Céline David ®,[2] Lena Berning ®,[2] Cathrin Nollmann ®,[1]
Thomas Lenz ®,[3] Kai Stühler ®,[3, 4] Thomas Heinzel ®,[1, *] and Björn Stork ®[2, †]

[1] *Solid State Physics Laboratory, Heinrich Heine University Düsseldorf, 40204 Düsseldorf, Germany*
[2] *Institute of Molecular Medicine I, Medical Faculty and University Hospital Düsseldorf,*
Heinrich Heine University, 40225 Düsseldorf, Germany
[3] *Molecular Proteomics Laboratory, Biological Medical Research Center,*
Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[4] *Institute of Molecular Medicine I, Proteome Research,*
Medical Faculty and University Hospital Düsseldorf,
Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
(Dated: June 22, 2024)

This supplement provides additional information regarding the characterization of the CNDs and the bPEI-CND conjugates, the uptake and intracellular distribution of the nanoparticles, the proteome and secretome analysis and the expression levels of p62, LC3 as well as cathepsin B.

---

[*] thomas.heinzel@hhu.de

[†] bjoern.stork@hhu.de

[1] D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. T. Doncheva, S. Pyysalo, P. Bork, L. J. Jensen, and C. von Mering, The string database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest, Nucleic Acids Research **51**, D638 (2023).
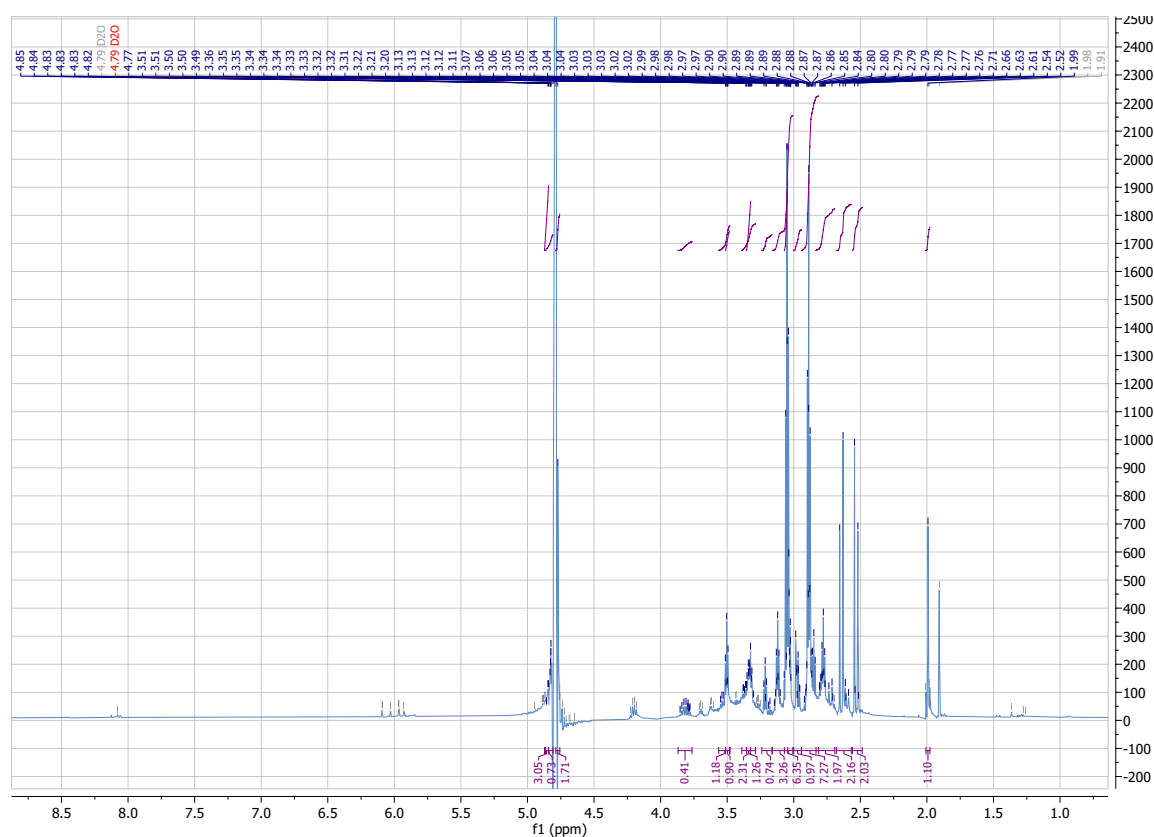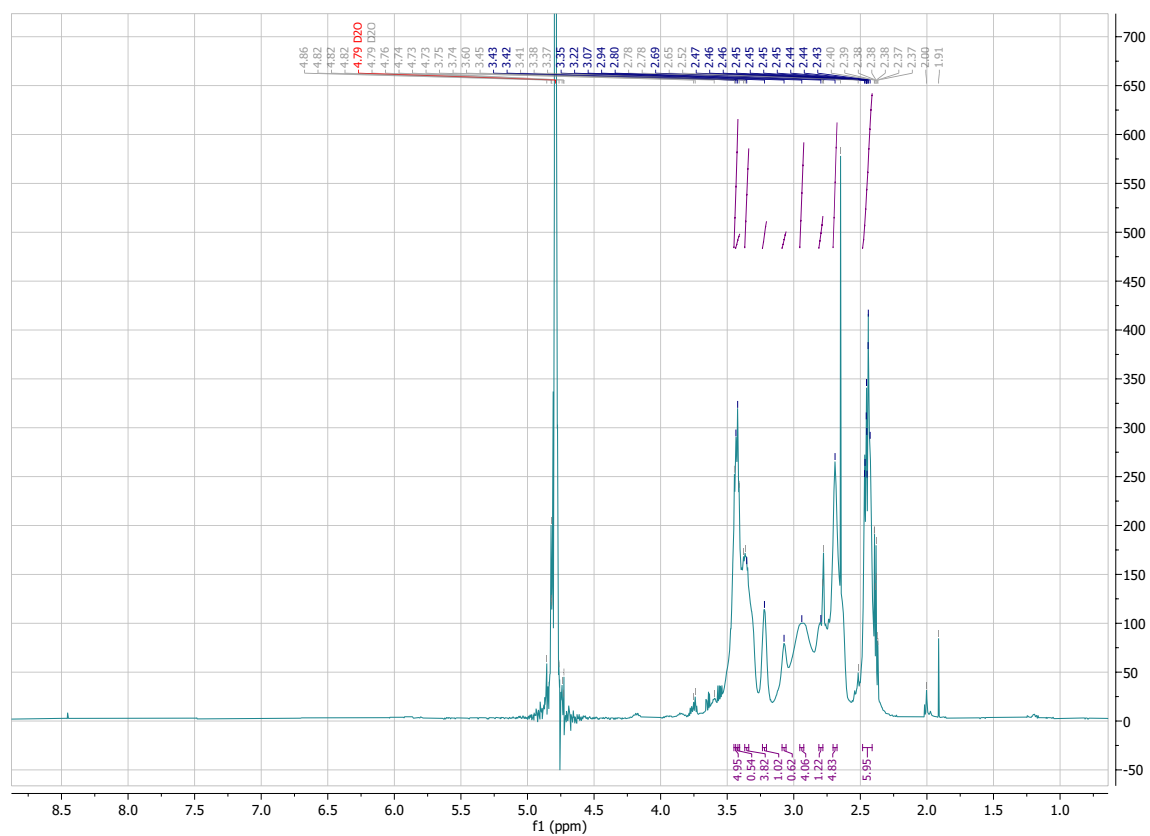
FIG. S1. ^{1}H NMR spectrum of CNDs. Peaks marked in blue, multiplets marked in purple. ^{1}H NMR (600 MHz, D$_2$O) δ[ppm]= 4.86 (s, 3H), 4.86 – 4.81 (m, 1H), 4.77 (s, 2H), 3.87 – 3.76 (m, 0H), 3.56 – 3.48 (m, 1H), 3.50 (s, 1H), 3.34 (s, 2H), 3.39 – 3.29 (m, 1H), 3.24 – 3.16 (m, 1H), 3.16 – 3.05 (m, 3H), 3.07 – 3.01 (m, 6H), 3.00 – 2.94 (m, 1H), 2.94 – 2.81 (m, 7H), 2.83 – 2.69 (m, 2H), 2.64 (d, J = 15.0 Hz, 2H), 2.53 (d, J = 15.0 Hz, 2H), 1.99 (s, 1H)

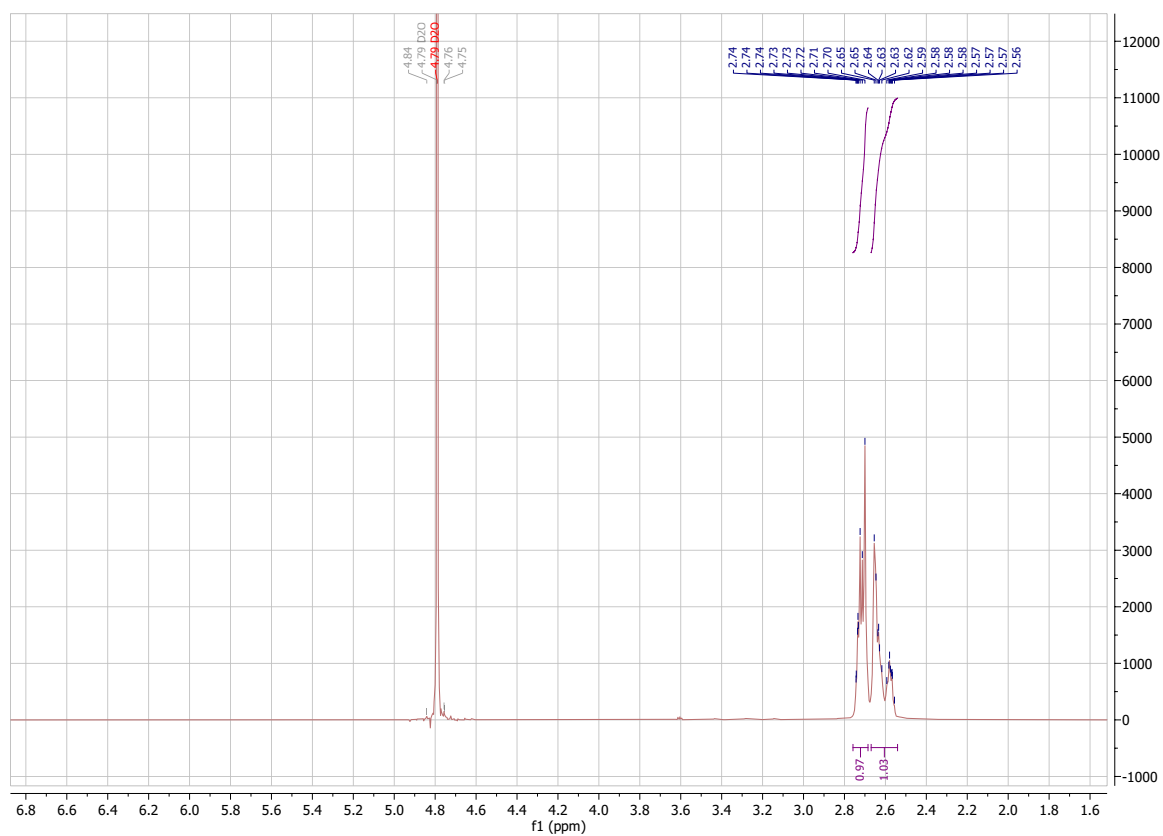FIG. S2. ^{1}H NMR spectrum of bPEI-CNDs. Peaks marked in blue, multiplets marked in purple. ^{1}H NMR (600 MHz, D$_2$O) δ[ppm]= 3.43 (s, 5H), 3.42 (s, 1H), 3.35 (s, 4H), 3.22 (s, 1H), 3.07 (s, 1H), 2.94 (s, 4H), 2.80 (s, 1H), 2.69 (s, 5H), 2.48 – 2.41 (m, 6H)

FIG. S3. ^{1}H NMR spectrum of bPEI. Peaks marked in blue, multiplets marked in purple. ^{1}H NMR (600 MHz, D$_2$O) δ[ppm]= 2.72 (dd, J = 13.5, 7.0 Hz, 1H), 2.67 – 2.54 (m, 1H)

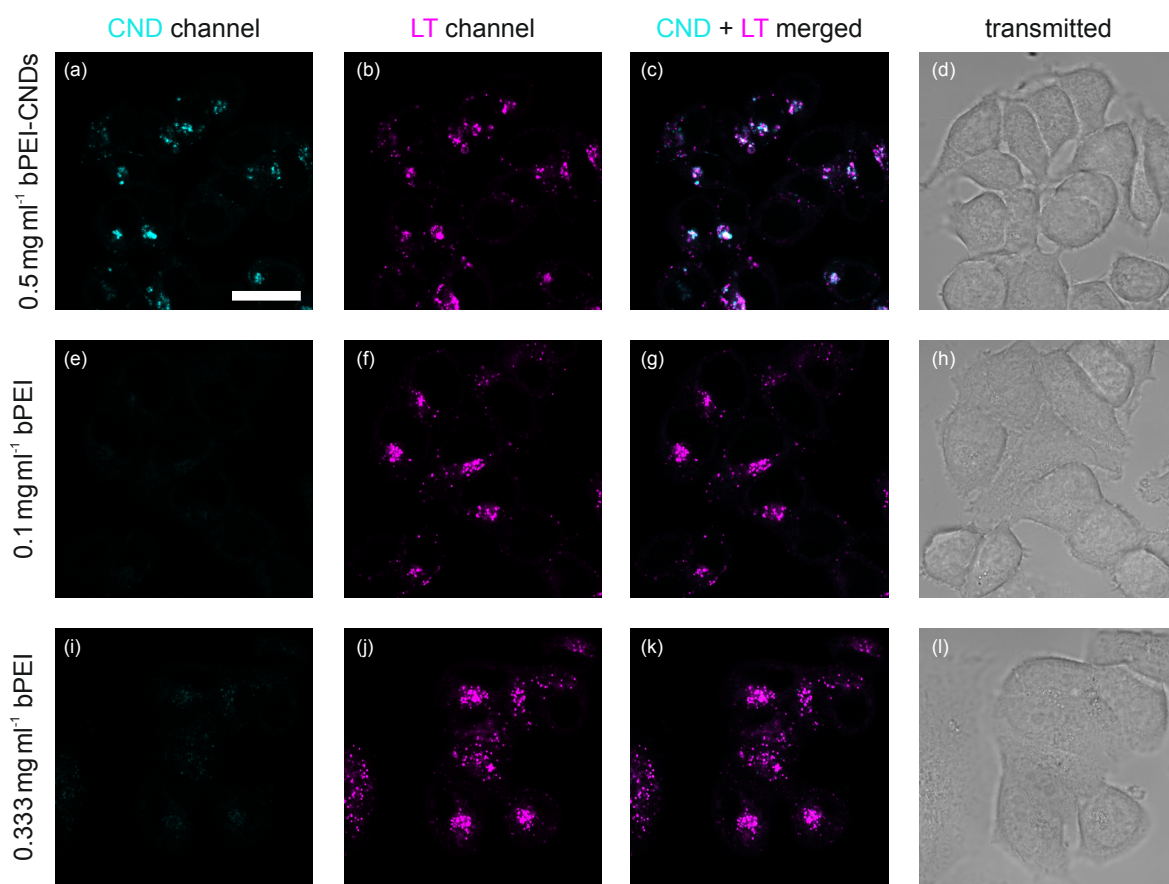| CND channel	LT channel	CND + LT merged	transmitted

FIG. S4. Confocal microscopy images of MCF-7 cells incubated with bPEI-CNDs or bPEI. Lysosomes are stained with lysotracker. The CND channel is shown in cyan, the lysotracker ("LT") channel in magenta. (a-d) MCF-7 cells incubated with $0.5\,\mathrm{mg\,ml^{-1}}$ bPEI-CNDs. (e-h) MCF-7 cells incubated with $0.1\,\mathrm{mg\,ml^{-1}}$ bPEI. (i-l) MCF-7 cells incubated with $0.333\,\mathrm{mg\,ml^{-1}}$ bPEI. The scale bar of $20\,\mathrm{\mu m}$ shown in (a) applies to all images.

```
close("*");
//programm written for .czi images. Images should all
be obtained with same measurement parameters.

//define directorys
inputdir = getDirectory("input");
dir = getDirectory("output");
dir2  = getDirectory("Zell-ROIs"); //in this directory, cell
ROIs (each cell is one ROI) should be saved in zip files
with the name: [title of image].czi_cells.zip
dir3  = getDirectory("Lysosome areas");

list = getFileList(inputdir);
for (p = 0; p < list.length; p++) {

//open image from the input directory and get infos
open(inputdir+list[p]);
title = File.nameWithoutExtension;
title2 = getInfo("image.filename");

imagedir = getInfo("image.directory");

//start ROI manager and open the previously defined
cell ROIs
run("ROI Manager...");
roiManager("Open", dir2+title+".czi_cells.zip");

n = getNumber("How many cell-ROIs are shown in the
ROI-Manager?", 10);


//start image analysis
for (i = 0; i < n; i++) {
//use a duplicate for analysis to not overwrite the
original image:
selectWindow(title2);
run("Duplicate...", "duplicate");
//Name of the duplicate will be Name-1.czi

selectWindow(title+"-1.czi");

//Start image processing with convoluted background
subtraction and intensity threshold setting. Make
image binary for further process. Intensity threshold
should be set once for all images.
Stack.setPosition(1,1,1);
run("Convoluted Background Subtraction",
"convolution=Median radius=10 slice");
setThreshold(10, 255);
run("Make Binary", "calculate only black");
run("Watershed", "slice");

//ROI manager will be opened to start with the analysis
of each cell separately
run("ROI Manager...");
roiManager("Show All");
roiManager("Delete");
```

```
roiManager("Show None");
roiManager("Open", dir2+title+".czi_cells.zip");

x=i+1;

selectWindow("ROI Manager");
roiManager("Select",i);
//lysosome channel should be the first image in the
stack.
selectWindow(title+"-1.czi");
Stack.setPosition(1,1,1);

//use particle analyzer to count particles and
determine size. Results should be saved (e.g. in excel
or txt table) an can be further analysed
run("Analyze Particles...", "size=0.01-Infinity
show=Outlines exclude clear summarize add slice");

//sav an image of the lysosomes of each cell
Stack.setPosition(1,1,1);
saveAs("Tiff", dir+title+"_Bild_lysosomen"+"_cell"+x);
close();


selectWindow(title+"-1.czi");
close();

//save lysosome ROIs
selectWindow("ROI Manager");
roiManager("Save",
dir+title+"lysosomes_cell"+x+"_ROI.zip");
run("Close");

//close all open windows for analysis of next image
selectWindow(title2);
close("\\Others");

run("ROI Manager...");
roiManager("Open", dir2+title+".czi_cells.zip");

}
run("ROI Manager...");
roiManager("Show All");
roiManager("Delete");
roiManager("Show None");
}
```

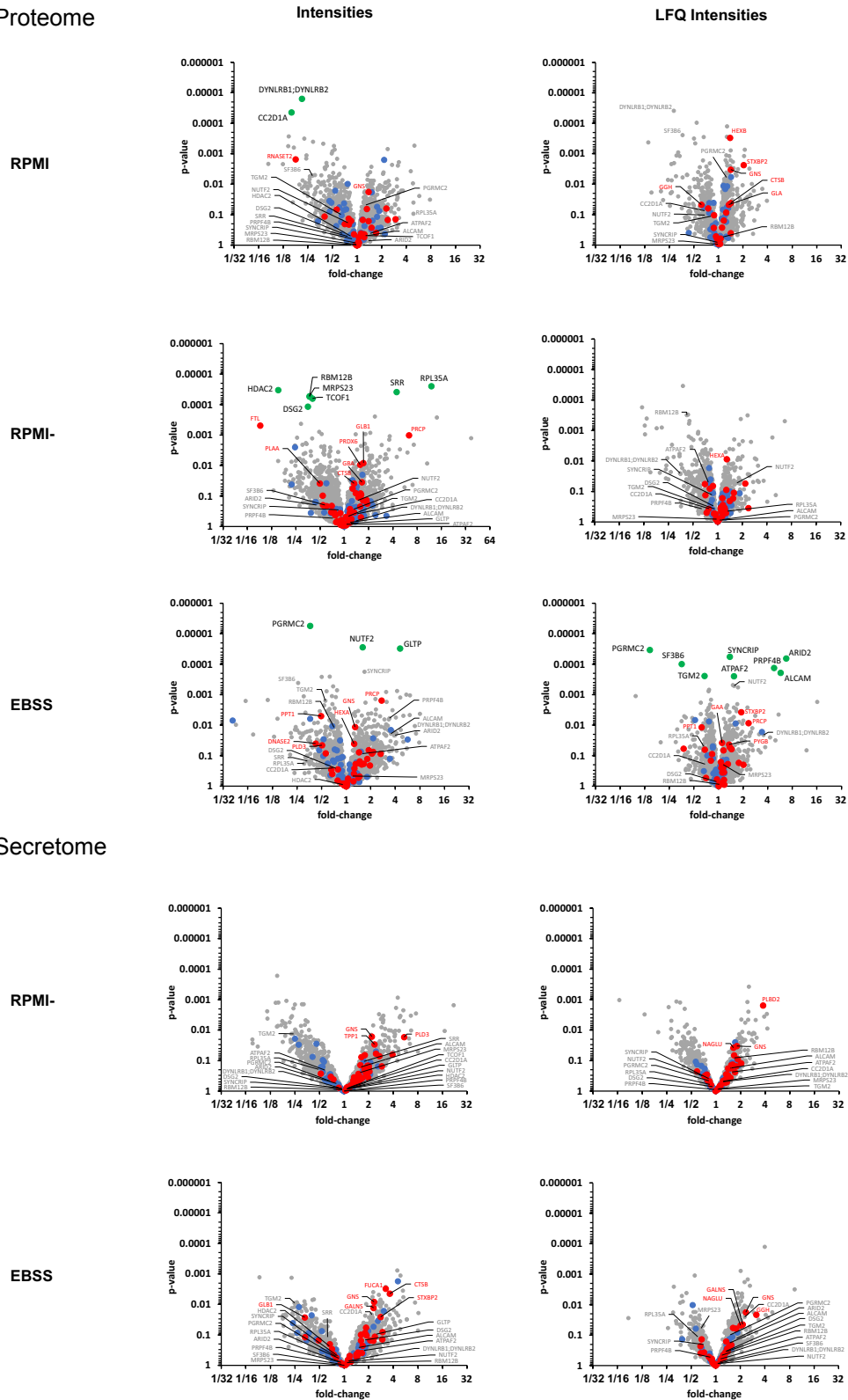FIG. S5. ImageJ Macro for lysosomal analysis

FIG. S6. Volcano plots based on intensity or LFQ intensity values (MaxQuant) for MS-based proteomics and secretomics. Analysis of differential protein abundance after CND treatment of MCF-7 cells was performed in different media (full medium "RPMI", basal medium RPMI1640 "RPMI-" or EBSS) as indicated. Proteins with a permutation-based FDR (SAM analysis) $< 5\,\%$ are indicated by green data points and labelled with their gene names in black for the respective conditions or in grey if this criterion was met for other conditions. Lysosomal lumenal and lysosomal membrane proteins are labelled with blue and red data points, respectively, and lumenal proteins with p-value < 0.05 for the respective conditions are labelled with red gene names.
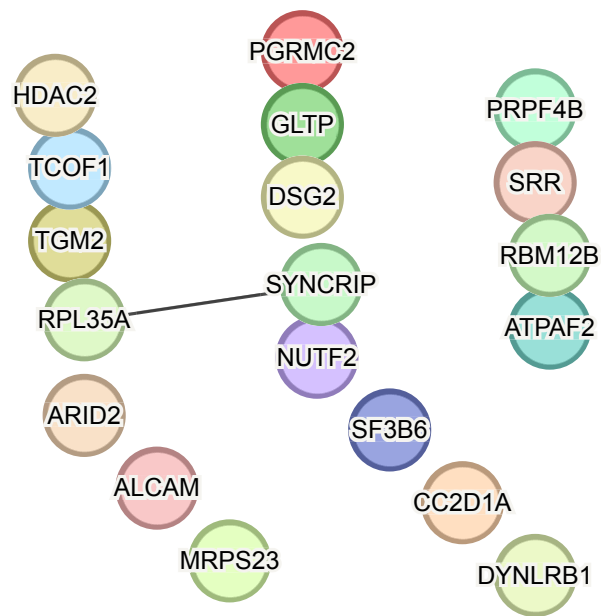
FIG. S7. Results of the functional enrichment analysis using the STRING database [1] for the significantly altered proteins from the proteome analysis. The black connection between *RPL35A* and *SYNCRIP* indicates co-expression of those two proteins. Analysis parameter: String Version 12.0, minimum required interaction score: 0.4. Network Stats: number of nodes: 19, number of edges: 1, average node degree: 0.105, avg. local clustering coefficient: 0.105, expected number of edges: 2, PPI enrichment p-value: 0.887
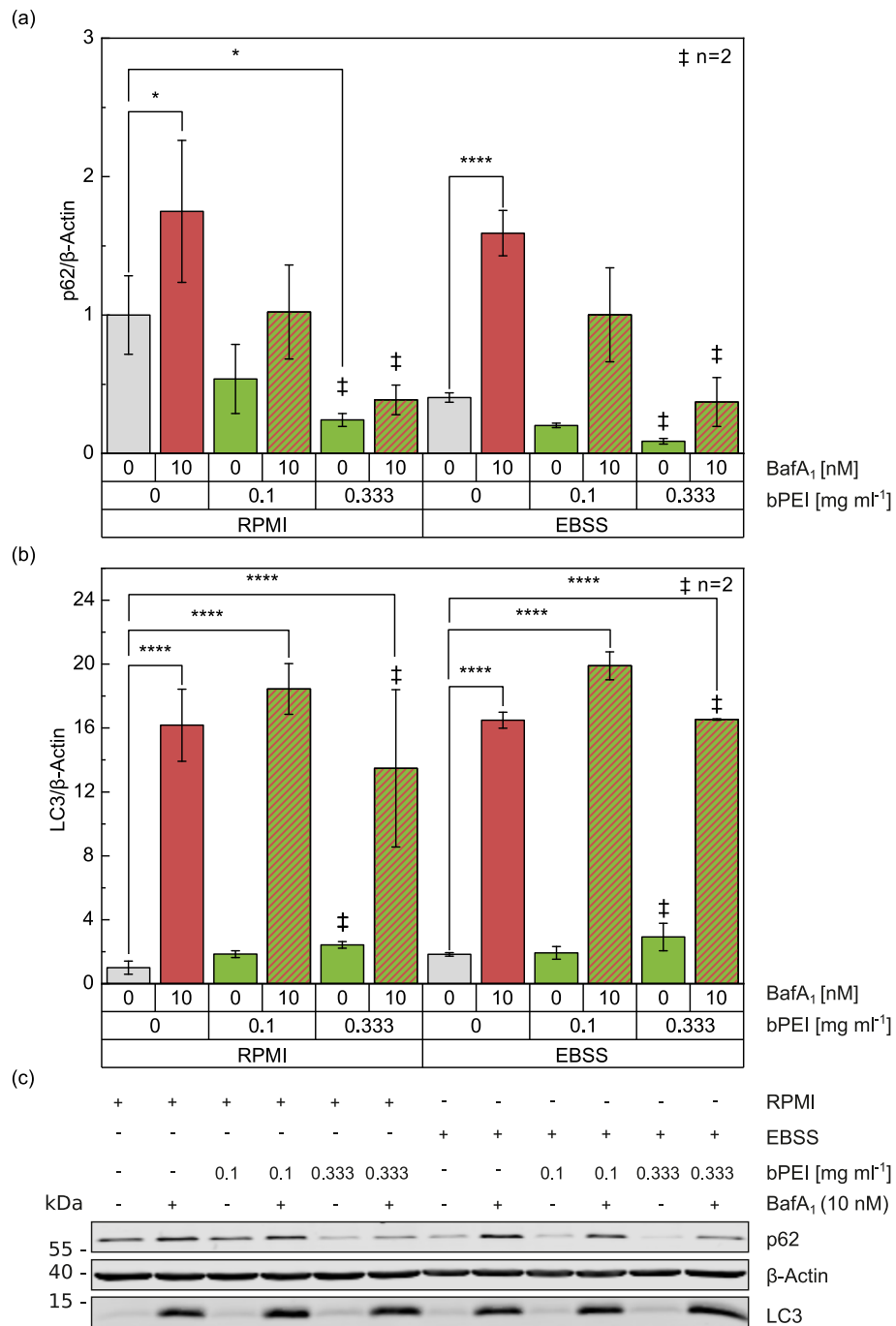
FIG. S8. (a) p62 and (b) LC3 levels of cells that are exposed to bPEI for 48 h and/or BafA$_1$ in full medium "RPMI" or starvation medium "EBSS" for 6 h. The levels were quantified and normalized to β-Actin. Data are represented as mean ± SD. A representativ immunoblot is shown in (c). p values were determined by three-way ANOVA with Bonferroni comparison. Only the significance values compared to the control samples "ct" in the same medium are shown. *p<0.05, ****p<0.0001.
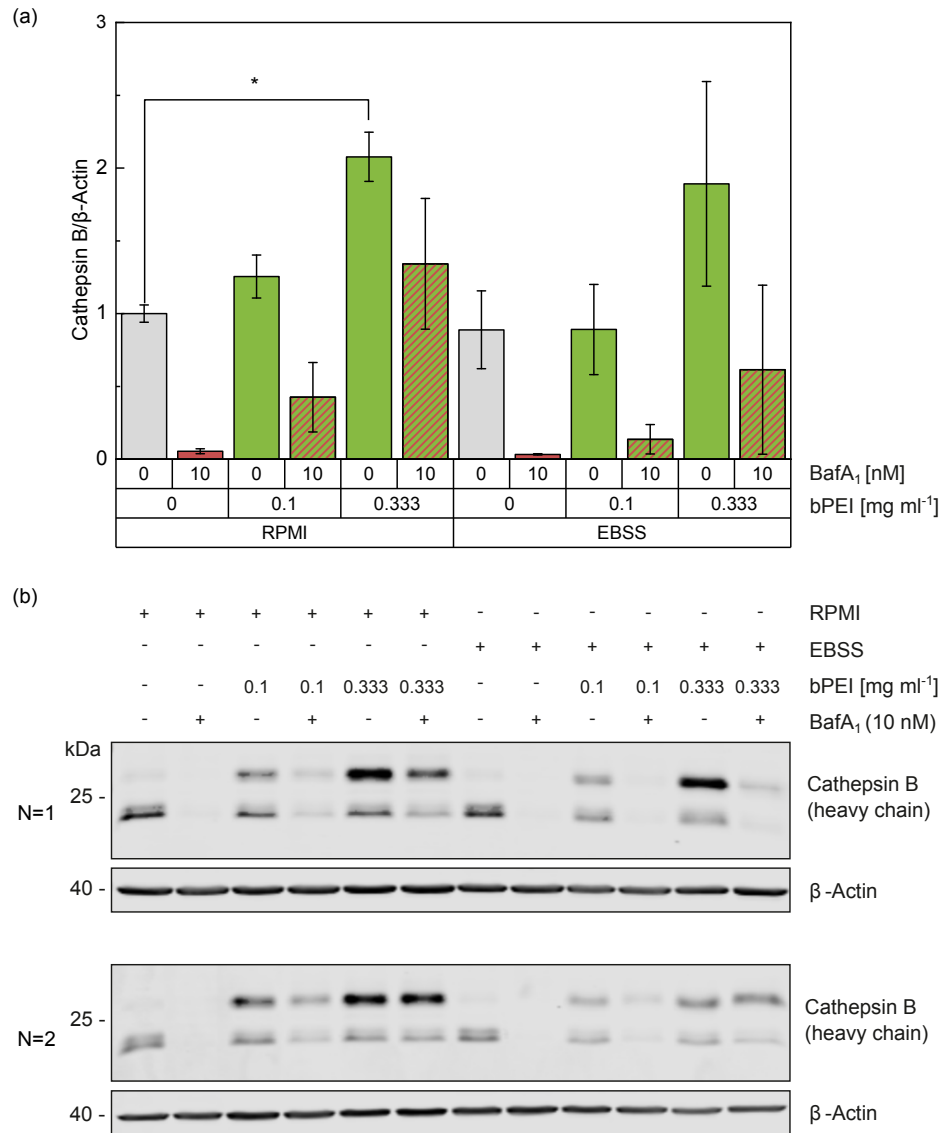
FIG. S9. (a) Cathepsin B levels of cells that are exposed to bPEI for 48 h and/or BafA$_1$ in full medium "RPMI" or starvation medium "EBSS" for 6 h. The levels were quantified and normalized to β-Actin. Data are represented as mean ± SD. A representativ immunoblot is shown in (b). p values were determined by three-way ANOVA with Bonferroni comparison. Only the significance values compared to the control samples "ct" in the same medium are shown. *p<0.05

5. Conclusion and outlook

This thesis investigates how AI algorithms such as t-SNE can contribute to the analysis of flow cytometry data from human hematopoietic cells and support clinical diagnostics. First, HSPCs were examined using a developed t-SNE protocol. The immunological phenotype of HSPCs was analyzed based on the expression of the markers CD34, CD38, CD123, CD45RA and PD-L1 and compared with the immunophenotype of blasts and LSCs in AML and MDS. Using the t-SNE map created from the common data set, the manually defined gates could be assigned to diagnostically relevant cell populations, two of which were predominantly leukemia-related. It was shown that this method can be applied to resolve fine substructures within a cell population, reflecting the subtle differences between the cells. For the selected markers it has been shown that the differences between HSPCs in comparison to blasts, and LSCs in AML and MDS are likely to be determined by the different proportions of certain cell types and different intensities of fluorescent markers within the cell types, rather than the exclusive presence of certain cell types. Furthermore, a method for classifying new samples based on the similarity comparison with t-SNE reference maps was proposed, using the pearson coefficient as a quantitative measure. This concept can also be applied to monitor the evolution of cell populations or patients undergoing therapy.

The uptake of CNDs in hematopoietic cells from healthy donors compared to patients with AML was examined as a parameter in the AI-supported evaluation. There were significant differences in the uptake between the two groups and between different cell types, which could be visualized concisely in a two-dimensional plot using t-SNE. While all cell types took up the CNDs, the $CD34^+$ and $CD33^+$ subsets of the AML samples showed a significantly reduced uptake. Having shown in principle, that the CNDs can selectively address specific cell subsets, an attempt was made to increase selective uptake by modifying the CNDs. For this purpose, the CNDs were modified with various monosaccharides and glycooligomers. No differences were observed in the uptake of CNDs between the different cell types as a function of the receptors, which are expressed according to the literature. On the other hand, differences were found in the uptake rate. CNDs that were functionalised with monosaccharides such as mannose or galactose showed a two- to three-fold increase in uptake compared to pristine CNDs and other conjugates. One preliminary explanation for the increased uptake of the monosaccharides could be the increased level of adsorption to the cell membrane through electrostatic interactions or sugar-mediated, cell type-independent selective binding.

Finally, the influence of CNDs on the lysosomes and the process of autophagy was investigated, as well as whether the CNDs are suitable as carriers for drug delivery into the lysosomes. The CNDs were shown to be inert as carriers and did not significantly affect lysosomal processes and autophagy as characterised by cathepsin B and L,the autophagy markers SQSTM1/p62 and LC3, and proteome analysis. As an example of a drug bPEI was coupled to the CNDs. It could be shown that bPEI was successfully delivered into lysosomes by the CNDs, confocal florescence microscopy revealing increased accumulation of bPEI-CNDs in lysosomes compared to pristine CNDs. In addition, it was

observed that the effects of free bPEI on the cells were attenuated by binding to the CNDs.

In summary, questions regarding the contribution of AI-assisted analysis of flow cytometry data of hematopoietic cells and clinical diagnostics have been successfully answered by the work this thesis is based on. The t-SNE supported analysis of hematopoietic cells could - by expanding to more markers - help to further characterize cell subsets on the basis of the rich substructure resolvable by t-SNE and to identify possible differences between healthy and diseased cells in order to derive potentially suitable therapeutic measures. By using more markers, which ideally exclude each other, a better separation of the cell populations within the t-SNE plot may be achieved [22]. This may be used to automate the gating process utilizing algorithms that, for example, recognize density-based clusters without specifying a number of clusters in advance. In order to further develop the classification of new samples using a t-SNE reference image, it is necessary that the reference image has a consistent and unchanging assignment of the cell types on the t-SNE map, consisting of as many samples as possible, so that subtypes of the disease to be diagnosed are adequately represented. Methods recently presented by D. Kobak, and P. Berens [23] as well as Poličar et al. [63] for embedding new samples into an existing t-SNE plot for single-cell transcriptomics data could be integrated into the algorithm to enable classification without any impact on the reference image once new samples are inserted. It is worth noting that the concepts presented are not limited to flow cytometry data from patients with AML and MDS, but can be applied to any multiparameter diagnostic flow cytometry data or to single-cell data in general. Other algorithms should also be tested for possible advantages in specific applications.
Regarding the selective uptake of the CNDs synthesized by the working group and their influence on cellular processes and use as a carrier in drug delivery systems relevant questions were answered and new areas of investigation have opened up driving the need for future work. The predominant localization in the lysosomes observed to date could be used for targeted drug delivery, for example to address lysosomal storage diseases or neurodegenerative diseases. It should be noted that the coupling of the compound to the CNDs may negatively impact the therapeutically effect as observed in the case of bPEI. The CNDs appear to be promising candidates for the development of a drug delivery system, as they have so far exposed characteristics as inert carriers regarding the lysosomal function and autophagy processes. Consequently, coupling the CNDs with corresponding drugs could be investigated. Organoids would be particularly suitable for this purpose, as they come closer to the complex arrangement and interaction of cells in the human body than the conventional two-dimensional cell culture. Selective uptake should also be investigated further and attempts should be made to reinforce the trends found.

Bibliography

[1] Z. Obermeyer and E. J. Emanuel, "Predicting the Future - Big Data, Machine Learning, and Clinical Medicine," *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016.

[2] K.-H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nature Biomedical Engineering*, vol. 2, no. 10, pp. 719–731, 2018.

[3] C. Caudai, A. Galizia, F. Geraci, L. Le Pera, V. Morea, E. Salerno, A. Via, and T. Colombo, "AI applications in functional genomics," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 5762–5790, 2021.

[4] J. Goecks, V. Jalili, L. M. Heiser, and J. W. Gray, "How Machine Learning Will Transform Biomedicine," *Cell*, vol. 181, no. 1, pp. 92–101, 2020.

[5] V. Kaul, S. Enslin, and S. A. Gross, "History of artificial intelligence in medicine," *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 807–812, 2020.

[6] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.

[7] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: Past, present and future," *Stroke and Vascular Neurology*, vol. 2, no. 4, pp. 230–243, 2017.

[8] A. Adan, G. Alizada, Y. Kiraz, Y. Baran, and A. Nalbant, "Flow cytometry: Basic principles and applications," *Critical Reviews in Biotechnology*, vol. 37, no. 2, pp. 163–176, 2017.

[9] M. L. Bisegna, I. Cordone, N. Peragine, M. L. Milani, S. Intoppa, P. de Fabritiis, M. Martelli, and M. S. De Propris, "Neoplastic bone marrow invasion:rapid exclusion of hematological disease by flow cytometric routine panels," *Blood Cells, Molecules and Diseases*, vol. 99, 2023.

[10] C. Duetz, S. Van Gassen, T. M. Westers, M. F. van Spronsen, C. Bachas, Y. Saeys, and A. A. van de Loosdrecht, "Computational flow cytometry as a diagnostic tool in suspected-myelodysplastic syndromes," *Cytometry Part A*, vol. 99, no. 8, pp. 814–824, 2021.

[11] C. Duetz, C. Bachas, T. M. Westers, and A. A. van de Loosdrecht, "Computational analysis of flow cytometry data in hematological malignancies: Future clinical practice?" *Current Opinion in Oncology*, vol. 32, no. 2, pp. 162–169, 2020.

[12] S. Cherian, B. D. Hedley, and M. Keeney, "Common flow cytometry pitfalls in diagnostic hematopathology," *Cytometry Part B: Clinical Cytometry*, vol. 96, no. 6, pp. 449–463, 2019.

[13] M. G. Della Porta and C. Picone, "Diagnostic utility of flow cytometry in myelodysplastic syndromes," *Mediterranean Journal of Hematology and Infectious Diseases*, vol. 9, 2017.

[14] M. S. Ahluwalia, P. K. Wallace, and D. M. Peereboom, "Flow cytometry as a diagnostic tool in lymphomatous or leukemic meningitis," *Cancer*, vol. 118, no. 7, pp. 1747–1753, 2012.

[15] J. P. Robinson, "Flow cytometry: Past and future," *BioTechniques*, vol. 72, no. 4, pp. 159–169, 2022.

[16] C. T. Mattern, F. S. Bracket, and B. J. Olson, "Determination of number and size of particles by electrical gating: Blood cells," *Journal of Applied Physiology*, vol. 10, no. 1, pp. 56–70, 1957.

[17] Y. Saeys, S. Van Gassen, and B. N. Lambrecht, "Computational flow cytometry: Helping to make sense of high-dimensional immunology data," *Nature Reviews Immunology*, vol. 16, no. 7, pp. 449–462, 2016.

[18] F. Mair, F. J. Hartmann, D. Mrdjen, V. Tosevski, C. Krieg, and B. Becher, "The end of gating? An introduction to automated analysis of high dimensional cytometry data," *European Journal of Immunology*, vol. 46, no. 1, pp. 34–43, 2016.

[19] P. Liu, S. Liu, Y. Fang, X. N. Xue, J. Zou, G. Tseng, and L. Konnikova, "Recent Advances in Computer-Assisted Algorithms for Cell Subtype Identification of Cytometry Data," *Frontiers in Cell and Developmental Biology*, vol. 8, 2020.

[20] L. M. Weber and M. D. Robinson, "Comparison of Clustering Methods for High-Dimensional Single-Cell Flow and Mass Cytometry Data," *Cytometry Part A*, vol. 89A, no. 12, pp. 1084–1096, 2016.

[21] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[22] S. T. Eshghi, A. Au-Yeung, C. Takahashi, C. R. Bolen, M. N. Nyachienga, S. P. Lear, C. Green, W. R. Mathews, and W. E. O'Gorman, "Quantitative Comparison of Conventional and t-SNE-guided Gating Analyses," *Frontiers in Immunology*, vol. 10, 2019.

[23] D. Kobak and P. Berens, "The art of using t-SNE for single-cell transcriptomics," *Nature Communications*, vol. 10, no. 1, 2019.

[24] Y.-H. Chien, K. K. Chan, T. Anderson, K. V. Kong, B. K. Ng, and K.-T. Yong, "Advanced Near-Infrared Light-Responsive Nanomaterials as Therapeutic Platforms for Cancer Therapy," *Advanced Therapeutics*, vol. 2, no. 3, 2019.

[25] N. Panwar, A. M. Soehartono, K. K. Chan, S. Zeng, G. Xu, J. Qu, P. Coquet, K.-T. Yong, and X. Chen, "Nanocarbons for Biology and Medicine: Sensing, Imaging, and Drug Delivery," *Chemical Reviews*, vol. 119, no. 16, pp. 9559–9656, 2019.

[26] Z. L. Wu, M. X. Gao, T. T. Wang, X. Y. Wan, L. L. Zheng, and C. Z. Huang, "A general quantitative pH sensor developed with dicyandiamide N-doped high quantum yield graphene quantum dots," *Nanoscale*, vol. 6, no. 7, pp. 3868–3874, 2014.

[27] H. Nie, M. Li, Q. Li, S. Liang, Y. Tan, L. Sheng, W. Shi, and S. X.-A. Zhang, "Carbon Dots with Continuously Tunable Full-Color Emission and Their Application in Ratiometric pH Sensing," *Chemistry of Materials*, vol. 26, no. 10, pp. 3104–3112, 2014.

[28] X. Qin, W. Lu, A. M. Asiri, A. O. Al-Youbi, and X. Sun, "Microwave-assisted rapid green synthesis of photoluminescent carbon nanodots from flour and their applications for sensitive and selective detection of mercury(ii) ions," *Sensors and Actuators B: Chemical*, vol. 184, pp. 156–162, 2013.

[29] X. Shan, L. Chai, J. Ma, Z. Qian, J. Chen, and H. Feng, "B-doped carbon quantum dots as a sensitive fluorescence probe for hydrogen peroxide and glucose detection," *Analyst*, vol. 139, no. 10, pp. 2322–2325, 2014.

[30] Q. Wang, X. Huang, Y. Long, X. Wang, H. Zhang, R. Zhu, L. Liang, P. Teng, and H. Zheng, "Hollow luminescent carbon dots for drug delivery," *Carbon*, vol. 59, pp. 192–199, 2013.

[31] J. Tang, B. Kong, H. Wu, M. Xu, Y. Wang, Y. Wang, D. Zhao, and G. Zheng, "Carbon Nanodots Featuring Efficient FRET for Real-Time Monitoring of Drug Delivery and Two-Photon Imaging," *Advanced Materials*, vol. 25, no. 45, pp. 6569–6574, 2013.

[32] Y. Liu, C. Zhao, A. Sabirsh, L. Ye, X. Wu, H. Lu, and J. Liu, "A Novel Graphene Quantum Dot-Based mRNA Delivery Platform," *ChemistryOpen*, vol. 10, no. 7, pp. 666–671, 2021.

[33] S. Mandal, S. R. Prasad, D. Mandal, and P. Das, "Bovine Serum Albumin Amplified Reactive Oxygen Species Generation from Anthrarufin-Derived Carbon Dot and Concomitant Nanoassembly for Combination Antibiotic-Photodynamic Therapy Application," *ACS Applied Materials & Interfaces*, vol. 11, no. 36, pp. 33 273–33 284, 2019.

[34] X. Dong, A. E. Bond, N. Pan, M. Coleman, Y. Tang, Y.-P. Sun, and L. Yang, "Synergistic photoactivated antimicrobial effects of carbon dots combined with dye photosensitizers," *International Journal of Nanomedicine*, vol. 13, pp. 8025–8035, 2018.

[35] J. S. Sidhu, Mayank, T. Pandiyan, N. Kaur, and N. Singh, "The Photochemical Degradation of Bacterial Cell Wall Using Penicillin-Based Carbon Dots: Weapons Against Multi-Drug Resistant (MDR) Strains," *ChemistrySelect*, vol. 2, no. 29, pp. 9277–9283, 2017.

[36] F. Zhao, W. Gu, J. Zhou, Q. Liu, and Y. Chong, "Solar-excited graphene quantum dots for bacterial inactivation via generation of reactive oxygen species," *Journal of Environmental Science and Health, Part C: Environmental Carcinogenesis & Ecotoxicology Reviews*, vol. 37, no. 2, pp. 67–80, 2019.

[37] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," *Metabolism: Clinical and Experimental*, vol. 69, no. S, S36–S40, 2017.

[38] C. Krittanawong, H. Zhang, Z. Wang, M. Aydar, and T. Kitai, "Artificial Intelligence in Precision Cardiovascular Medicine," *Journal of the American College of Cardiology*, vol. 69, no. 21, pp. 2657–2664, 2017.

[39] Y. Jiang, X. Li, H. Luo, S. Yin, and O. Kaynak, "Quo vadis artificial intelligence?" *Discover Artificial Intelligence*, vol. 2, no. 1, p. 4, 2022.

[40] S. Russell and P. Norvig: *Artificial Intelligence, Global Edition. A Modern Approach*. Pearson Deutschland, 2021.

[41] M. Shehab, L. Abualigah, Q. Shambour, M. A. Abu-Hashem, M. K. Y. Shambour, A. I. Alsalibi, and A. H. Gandomi, "Machine learning in medical applications: A review of state-of-the-art methods," *Computers in Biology and Medicine*, vol. 145, 2022.

[42] S. J. MacEachern and N. D. Forkert, "Machine learning for precision medicine," *Genome*, vol. 64, no. 4, pp. 416–425, 2021.

[43] A. Barragan-Montero, U. Javaid, G. Valdes, *et al.*, "Artificial intelligence and machine learning for medical imaging: A technology review," *Physica Medica: European Journal of Medical Physics*, vol. 83, pp. 242–256, 2021.

[44] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Computer Science Review*, vol. 40, 2021.

[45] C. L. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[46] O. Kokshagina, P. Le Masson, and J. Luo, "Beyond the data fads: Impact of big data on contemporary innovation and technology management," *Technovation*, vol. 134, p. 103 026, 2024.

[47] J. Luo, "Data-Driven Innovation: What Is It?" *IEEE Transactions on Engineering Management*, vol. 70, no. 2, pp. 784–790, 2023.

[48] S. Abe: "Feature selection and extraction," *Support Vector Machines for Pattern Classification*. London: Springer London, 2010, pp. 331–341.

[49] S. Marukatat, "Tutorial on PCA and approximate PCA and approximate kernel PCA," *Artificial Intelligence Review*, vol. 56, no. 6, pp. 5445–5477, 2023.

[50] A. Thakkar, N. Kikani, and R. Geddam, "Fusion of linear and non-linear dimensionality reduction techniques for feature reduction in LSTM-based Intrusion Detection System," *Applied Soft Computing*, vol. 154, 2024.

[51] Y. F. Wang, H. Y. Huang, C. Rudin, and Y. Shaposhnik, "Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization," *Journal of Machine Learning Research*, vol. 22, 2021.

[52] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9, pp. 2812–2831, 2014.

[53] K. Pearson, "On lines and planes of closest fit to points in space," *Philosophical Magazine*, 1901.

[54] H. Hotelling: "The generalization of student's ratio," *Breakthroughs in Statistics: Foundations and Basic Theory*, S. Kotz and N. L. Johnson, Eds. New York, NY: Springer New York, 1992, pp. 54–65.

[55] B. Ghojogh, M. Samad, S. Mashhadi, T. Kapoor, W. Ali, F. Karray, and M. Crowley, "Feature selection and feature extraction in pattern analysis: A literature review," *arXiv preprint arXiv:1905.02845*, 2019.

[56] G. E. Hinton and S. T. Roweis: "Stochastic neighbor embedding," *Proceedings of the 15th International Conference on Neural Information Processing Systems*, Cambridge, MA, USA: MIT Press, 2002, pp. 857–864.

[57] M. Wattenberg, F. Viégas, and I. Johnson, "How to Use t-SNE Effectively," *Distill*, 2016.

[58] R. Gove, L. Cadalzo, N. Leiby, J. M. Singer, and A. Zaitzeff, "New guidance for using t-SNE: Alternative defaults, hyperparameter selection automation, and comparative evaluation," *Visual Informatics*, vol. 6, no. 2, pp. 87–97, 2022.

[59] A. C. Belkina, C. O. Ciccolella, R. Anno, R. Halpert, J. Spidlen, and J. E. Snyder-Cappione, "Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets," *Nature Communications*, vol. 10, 2019.

[60] Y. Cao and L. Wang: *Automatic Selection of t-SNE Perplexity*, 2017. arXiv: 1708.03229 [cs.AI].

[61] C. P. Roca, O. T. Burton, J. Neumann, S. Tareen, C. E. Whyte, V. Gergelits, R. V. Veiga, S. Humblet-Baron, and A. Liston, "A cross entropy test allows quantitative statistical comparison of t-SNE and UMAP representations," *Cell Reports Methods*, vol. 3, no. 1, 2023.

[62] D. Kobak and G. C. Linderman, "Initialization is critical for preserving global data structure in both t-SNE and UMAP," *Nature Biotechnology*, vol. 39, no. 2, pp. 156–157, 2021.

[63] P. G. Poličar, M. Stražar, and B. Zupan, "Embedding to reference t-SNE space addresses batch effects in single-cell classification," *Machine Learning*, vol. 112, no. 2, pp. 721–740, 2023.

[64] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding, and C.-T. Lin, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, 2017.

[65] A. E. Ezugwu, A. M. Ikotun, O. O. Oyelade, L. Abualigah, J. O. Agushaka, C. I. Eke, and A. A. Akinyelu, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, 2022.

[66] P.-N. Tan, M. Steinbach, and V. Kumar: *Introduction to data mining*. Pearson Education India, 2016.

[67] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, SI, pp. 651–666, 2010.

[68] Y. Liu, X. Wu, and Y. Shen, "Automatic clustering using genetic algorithms," *Applied Mathematics and Computation*, vol. 218, no. 4, pp. 1267–1279, 2011.

[69]   Amisha, P. Malik, M. Pathania, and V. K. Rathaur, "Overview of artificial intelligence in medicine," *Journal of Family Medicine and Primary Care*, vol. 8, no. 7, pp. 2328–2331, 2019.

[70]   J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[71]   N. A. Obuchowski, "Receiver operating characteristic curves and their use in radiology," *Radiology*, vol. 229, no. 1, pp. 3–8, 2003.

[72]   A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[73]   L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, 2017.

[74]   D. Shen, G. Wu, and H.-I. Suk: "Deep Learning in Medical Image Analysis," *Annual Review of Biomedical Engineering*, ser. Annual Review of Biomedical Engineering, M. Yarmush, Ed., vol. 19, 2017, pp. 221–248.

[75]   D. Ueda, A. Shimazaki, and Y. Miki, "Technical and clinical overview of deep learning in radiology," *Japanese Journal of Radiology*, vol. 37, no. 1, pp. 15–33, 2019.

[76]   C. Huang, E. A. Clayton, L. Matyunina V, L. D. McDonald, B. B. Benigno, F. Vannberg, and J. F. McDonald, "Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy," *Scientific Reports*, vol. 8, 2018.

[77]   F. Michor and K. Polyak, "The Origins and Implications of Intratumor Heterogeneity," *Cancer Prevention Research*, vol. 3, no. 11, pp. 1361–1364, 2010.

[78]   K. Cyll, E. Ersvaer, L. Vlatkovic, *et al.*, "Tumour heterogeneity poses a significant challenge to cancer biomarker research," *British Journal of Cancer*, vol. 117, no. 3, pp. 367–375, 2017.

[79]   K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology," *Nature Reviews Clinical Oncology*, vol. 16, no. 11, pp. 703–715, 2019.

[80]   B. van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi: "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 286–289.

[81]   P. Lakhani and B. Sundaram, "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

[82]   X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers: *ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases*, 2017. arXiv: 1705.02315 [cs.CV].

[83] P. Rajpurkar, J. Irvin, K. Zhu, *et al.*: *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*, 2017. arXiv: `1711.05225` `[cs.CV]`.

[84] T. Wang, Y. Lei, Y. Fu, J. F. Wynne, W. J. Curran, T. Liu, and X. Yang, "A review on medical imaging synthesis using deep learning and its clinical applications," *Journal of Applied Clinical Medical Physics*, vol. 22, no. 1, pp. 11–36, 2021.

[85] H. Seo, M. B. Khuzani, V. Vasudevan, C. Huang, H. Ren, R. Xiao, X. Jia, and L. Xing, "Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications," *Medical Physics*, vol. 47, no. 5, E148–E167, 2020.

[86] M. H. S. Segler, T. Kogej, C. Tyrchan, and M. P. Waller, "Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks," *ACS Central Science*, vol. 4, no. 1, pp. 120–131, 2018.

[87] K.-K. Mak and M. R. Pichika, "Artificial intelligence in drug development: Present status and future prospects," *Drug Discovery Today*, vol. 24, no. 3, pp. 773–780, 2019.

[88] J. P. Hughes, S. Rees, S. B. Kalindjian, and K. L. Philpott, "Principles of early drug discovery," *British Journal of Pharmacology*, vol. 162, no. 6, pp. 1239–1249, 2011.

[89] A. A. Kalinin, G. A. Higgins, N. Reamaroon, S. Soroushmehr, A. Allyn-Feuer, I. D. Dinov, K. Najarian, and B. D. Athey, "Deep learning in pharmacogenomics: From gene regulation to patient stratification," *Pharmacogenomics*, vol. 19, no. 7, pp. 629–650, 2018.

[90] M. Cheung, J. J. Campbell, L. Whitby, R. J. Thomas, J. Braybrook, and J. Petzing, "Current trends in flow cytometry automated data analysis software," *Cytometry Part A*, vol. 99, no. 10, pp. 1007–1021, 2021.

[91] J. Quinn, P. W. Fisher, R. J. Capocasale, R. Achuthanandam, M. Kam, P. J. Bugelski, and L. Hrebien, "A statistical pattern recognition approach for determining cellular viability and lineage phenotype in cultured cells and murine bone marrow," *Cytometry Part A*, vol. 71A, no. 8, pp. 612–624, 2007.

[92] K. Lo, R. R. Brinkman, and R. Gottardo, "Automated gating of flow cytometry data via robust model-based clustering," *Cytometry Part A*, vol. 73A, no. 4, pp. 321–332, 2008.

[93] G. Finak, A. Bashashati, R. Brinkman, and R. Gottardo, "Merging mixture components for cell population identification in flow cytometry," *Advances in bioinformatics*, vol. 2009, no. 1, p. 247 646, 2009.

[94] S. Pyne, X. L. Hu, K. Wang, *et al.*, "Automated high-dimensional flow cytometric data analysis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 21, pp. 8519–8524, 2009.

[95] I. P. Sugar and S. C. Sealfon, "Misty mountain clustering: Application to fast unsupervised flow cytometry gating," *BMC Bioinformatics*, vol. 11, 2010.

[96] N. Aghaeepour, R. Nikolic, H. H. Hoos, and R. R. Brinkman, "Rapid Cell Population Identification in Flow Cytometry Data," *Cytometry Part A*, vol. 79A, no. 1, pp. 6–13, 2011.

[97] N. Aghaeepour, G. Finak, H. Hoos, T. R. Mosmann, R. Brinkman, R. Gottardo, R. H. Scheuermann, C. A. P. C. Flow, and D. Consortium, "Critical assessment of automated flow cytometry data analysis techniques," *Nature Methods*, vol. 10, no. 3, pp. 228–238, 2013.

[98] P. Qiu, "Computational prediction of manually gated rare cells in flow cytometry data," *Cytometry Part A*, vol. 87A, no. 7, pp. 594–602, 2015.

[99] N. Aghaeepour, P. Chattopadhyay, M. Chikina, *et al.*, "A benchmark for evaluation of algorithms for identification of cellular correlates of clinical outcomes," *Cytometry Part A*, vol. 89A, no. 1, pp. 16–21, 2016.

[100] C. P. Verschoor, A. Lelic, J. L. Bramson, and D. M. E. Bowdish, "An introduction to automated flow cytometry gating tools and their implementation," *Frontiers in Immunology*, vol. 6, 2015.

[101] Z. C. Hu, S. Bhattacharya, and A. J. Butte, "Application of Machine Learning for Cytometry Data," *Frontiers in Immunology*, vol. 12, 2022.

[102] M. C. Béné, F. Lacombe, and A. Porwit, "Unsupervised flow cytometry analysis in hematological malignancies: A new paradigm," *International Journal of Laboratory Hematology*, vol. 43, no. 1, pp. 54–64, 2021.

[103] P. Q. Zhong, M. Z. Hong, H. Y. He, J. Zhang, Y. M. Chen, Z. G. Wang, P. S. Chen, and J. Ouyang, "Diagnosis of Acute Leukemia by Multiparameter Flow Cytometry with the Assistance of Artificial Intelligence," *Diagnostics*, vol. 12, no. 4, 2022.

[104] C.-M. Aanei, R. Veyrat-Masson, L. Rigollet, J. Stagnara, E. T. Tardy, E. Daguenet, D. Guyotat, and L. C. Catafal, "Advanced Flow Cytometry Analysis Algorithms for Optimizing the Detection of "Different From Normal" Immunophenotypes in Acute Myeloid Blasts," *Frontiers in Cell and Developmental Biology*, vol. 9, 2021.

[105] F.-M. Cheng, S.-C. Lo, C.-C. Lin, W.-J. Lo, S.-Y. Chien, T.-H. Sun, and K.-C. Hsu, "Deep learning assists in acute leukemia detection and cell classification via flow cytometry using the acute leukemia orientation tube," *Scientific Reports*, vol. 14, no. 1, 2024.

[106] V. Clichet, D. Lebon, N. Chapuis, J. Zhu, V. Bardet, J.-P. Marolleau, L. Garcon, A. Caulier, and T. Boyer, "Artificial intelligence to empower diagnosis of myelodysplastic syndromes by multiparametric flow cytometry," *Haematologica*, vol. 108, no. 9, pp. 2435–2443, 2023.

[107] J. P. Vial, N. Lechevalier, F. Lacombe, P.-Y. Dumas, A. Bidet, T. Leguay, F. Vergez, A. Pigneux, and M. C. Bene, "Unsupervised Flow Cytometry Analysis Allows for an Accurate Identification of Minimal Residual Disease Assessment in Acute Myeloid Leukemia," *Cancers*, vol. 13, no. 4, 2021.

[108] M. E. Salama, G. E. Otteson, J. J. Camp, J. N. Seheult, D. Jevremovic, D. R. Holmes III, H. Olteanu, and M. Shi, "Artificial Intelligence Enhances Diagnostic Flow Cytometry Workflow in the Detection of Minimal Residual Disease of Chronic Lymphocytic Leukemia," *Cancers*, vol. 14, no. 10, 2022.

[109] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, "Electric Field Effect in Atomically Thin Carbon Films," *Science*, vol. 306, no. 5696, pp. 666–669, 2004.

[110] N. Mermin and H. Wagner, "Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic heisenberg models," *Physical Review Letters*, vol. 17, no. 22, pp. 1133–1136, 1966.

[111] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, "The electronic properties of graphene," *Review of modern physics*, vol. 81, no. 1, pp. 109–163, 2009. DOI: https://doi.org/10.1103/RevModPhys.81.109.

[112] X. T. Zheng, A. Ananthanarayanan, K. Q. Luo, and P. Chen, "Glowing Graphene Quantum Dots and Carbon Dots: Properties, Syntheses, and Biological Applications," *Small*, vol. 11, no. 14, pp. 1620–1636, 2015.

[113] X. Zhang, C. Wei, Y. Li, and D. Yu, "Shining luminescent graphene quantum dots: Synthesis, physicochemical properties, and biomedical applications," *TrAC Trends in Analytical Chemistry*, vol. 116, pp. 109–121, 2019.

[114] G. Eda, Y.-Y. Lin, C. Mattevi, H. Yamaguchi, H.-A. Chen, I.-S. Chen, C.-W. Chen, and M. Chhowalla, "Blue Photoluminescence from Chemically Derived Graphene Oxide," *Advanced Materials*, vol. 22, no. 4, pp. 505–509, 2010.

[115] S. Kim, S. W. Hwang, M.-K. Kim, *et al.*, "Anomalous behaviors of Visible Luminescence from Graphene Quantum Dots: Interplay between Size and Shape," *ACS Nano*, vol. 6, no. 9, pp. 8203–8208, 2012.

[116] K. Ritter and L. J.W., "The influence of edge structure on the electronic properties of graphene quantum dots and nanoribbons," *Nature Materials*, vol. 8, no. 3, pp. 235–242, 2009.

[117] S. H. Jin, D. H. Kim, G. H. Jun, S. H. Hong, and S. Jeon, "Tuning the Photoluminescence of Graphene Quantum Dots through the Charge transfer Effect of Functional Groups," *ACS Nano*, vol. 7, no. 2, pp. 1239–1245, 2013.

[118] S. Zhu, Y. Song, X. Zhao, J. Shao, J. Zhang, and B. Yang, "The photoluminescence mechanism in carbon dots (graphene quantum dots, carbon nanodots, and polymer dots): Current state and future perspective," *Nano Research*, vol. 8, no. 2, pp. 355–381, 2015.

[119] Z. L. Wu, M. X. Gao, T. T. Wang, X. Y. Wan, L. L. Zhenga, and C. Z. Huang, "A general quantitative pH sensor developed with dicyandiamide N-doped high quatum yield graphene quatum dots," *Nanoscale*, vol. 6, no. 7, pp. 3868–3874, 2014.

[120] S. Zhu, J. Zhang, C. Qiao, *et al.*, "Strongly green-photoluminescent graphene quantum dots for bioimaging applications," *Chemical Communications*, vol. 47, no. 24, pp. 6858–6860, 2011.

[121] D. Pan, J. Zhang, Z. Li, and M. Wu, "Hydrothermal Route for Cutting Graphene Sheets into Blue-Luminescent Graphene Quantum Dots," *Advanced Materials*, vol. 22, no. 6, pp. 734–738, 2010.

[122] J. Lu, J.-x. Yang, J. Wang, A. Lim, S. Wang, and K. P. Loh, "One-Pot Synthesis of Fluorescent Carbon Nanoribbons, Nanoparticles, and Graphene by the Exfoliation of Graphite in Ionic Liquids," *ACS Nano*, vol. 3, no. 8, pp. 2367–2375, 2009.

[123] L. Zheng, Y. Chi, Y. Dong, J. Lin, and B. Wang, "Electrochemiluminescence of Water-Soluble Carbon Nanocrystals Released Electrochemically from Graphite," *Journal of the American Chemical Society*, vol. 131, no. 13, pp. 4564–4565, 2009.

[124] L. Lin and S. Zhang, "Creating high yield water soluble luminescent graphene quantum dots via exfoliating and disintegrating carbon nanotubes and graphite flakes," *Chemical Communications*, vol. 48, no. 82, pp. 10 177–10 179, 2012.

[125] M. Bottini, C. Balasubramanian, M. Dawson, A. Bergamaschi, S. Bellucci, and T. Mustelin, "Isolation and characterization of fluorescent nanoparticles from pristine and oxidized electric arc-produced single-walled carbon nanotubes," *Journal of Physical Chemistry B*, vol. 110, no. 2, pp. 831–836, 2006.

[126] Y.-P. Sun, B. Zhou, Y. Lin, *et al.*, "Quantum-sized carbon dots for bright and colorful photoluminescence," *Journal of the American Chemical Society*, vol. 128, no. 24, pp. 7756–7757, 2006.

[127] J. Lee, K. Kim, W. I. Park, *et al.*, "Uniform Graphene Quantum Dots Patterned from Self-Assembled Silica Nanodots," *Nano Letters*, vol. 12, no. 12, pp. 6078–6083, 2012.

[128] L. Fan, M. Zhu, X. Lee, R. Zhang, K. Wang, J. Wei, M. Zhong, D. Wu, and H. Zhu, "Direct Synthesis of Graphene Quantum Dots by Chemical Vapor Deposition," *Particle & Particle Systems Characterization*, vol. 30, no. 9, pp. 764–769, 2013.

[129] B. Wang and S. Lu, "The light of carbon dots: From mechanism to applications," *Matter*, vol. 5, no. 1, pp. 110–149, 2022.

[130] Z.-C. Yang, M. Wang, A. M. Yong, S. Y. Wong, X.-H. Zhang, H. Tan, A. Y. Chang, X. Li, and J. Wang, "Intrinsically fluorescent carbon dots with tunable emission derived from hydrothermal treatment of glucose in the presence of monopotassium phosphate," *Chemical Communications*, vol. 47, no. 42, pp. 11 615–11 617, 2011.

[131] H. Zhu, X. Wang, Y. Li, Z. Wang, F. Yang, and X. Yang, "Microwave synthesis of fluorescent carbon nanoparticles with electrochemiluminescence properties," *Chemical Communications*, no. 34, pp. 5118–5120, 2009.

[132] A. B. Bourlinos, A. Stassinopoulos, D. Anglos, R. Zboril, M. Karakassides, and E. P. Giannelis, "Surface functionalized carbogenic quantum dots," *Small*, vol. 4, no. 4, pp. 455–458, 2008.

[133] H. Peng and J. Travas-Sejdic, "Simple Aqueous Solution Route to Luminescent Carbogenic Dots from Carbohydrates," *Chemistry of Materials*, vol. 21, no. 23, pp. 5563–5565, 2009.

[134] J. Zong, Y. Zhu, X. Yang, J. Shen, and C. Li, "Synthesis of photoluminescent carbogenic dots using mesoporous silica spheres as nanoreactors," *Chemical Communications*, vol. 47, no. 2, pp. 764–766, 2011.

[135] F. Yan, Y. Jiang, X. Sun, Z. Bai, Y. Zhang, and X. Zhou, "Surface modification and chemical functionalization of carbon dots: A review," *Microchimica Acta*, vol. 185, no. 9, 2018.

[136] B. Wang, H. Cai, G. I. N. Waterhouse, X. Qu, B. Yang, and S. Lu, "Carbon Dots in Bioimaging, Biosensing and Therapeutics: A Comprehensive Review," *Small Science*, vol. 2, no. 6, 2022.

[137] W. Dong, S. Zhou, Y. Dong, J. Wang, X. Ge, and L. Sui, "The preparation of ethylenediamine-modified fluorescent carbon dots and their use in imaging of cells," *Luminescence*, vol. 30, no. 6, pp. 867–871, 2015.

[138] W. Kasprzyk, P. P. Romanczyk, J. Feldmann, J. K. Stolarczyk, and T. Swiergosz, "The role of molecular fluorophores in the photoluminescence of carbon dots derived from citric acid: Current state-of-the-art and future perspectives," *Nanoscale*, vol. 14, no. 39, pp. 14 368–14 384, 2022.

[139] L. Wang, S.-J. Zhu, H.-Y. Wang, *et al.*, "Common Origin of Green Luminescence in Carbon Nanodots and Graphene Quantum Dots," *ACS Nano*, vol. 8, no. 3, pp. 2541–2547, 2014.

[140] S. Zhu, J. Shao, Y. Song, X. Zhao, J. Du, L. Wang, H. Wang, K. Zhang, J. Zhang, and B. Yang, "Investigating the surface state of graphene quantum dots," *Nanoscale*, vol. 7, no. 17, pp. 7927–7933, 2015.

[141] S. Zhu, L. Wang, N. Zhou, X. Zhao, Y. Song, S. Maharjan, J. Zhang, L. Lu, H. Wang, and B. Yang, "The crosslink enhanced emission (CEE) in non-conjugated polymer dots: from the photoluminescence mechanism to the cellular uptake mechanism and internalization," *Chemical Communications*, vol. 50, no. 89, pp. 13 845–13 848, 2014.

[142] J. Sun, J. Yu, Z. Jiang, Z. Zhao, and Y. Xia, "Fluorescent Carbonized Polymer Dots Prepared from Sodium Alginate Based on the CEE Effect," *ACS Omega*, vol. 5, no. 42, pp. 27 514–27 521, 2020.

[143] M. Fu, F. Ehrat, Y. Wang, K. Z. Milowska, C. Reckmeier, A. L. Rogach, J. K. Stolarczyk, A. S. Urban, and J. Feldmann, "Carbon Dots: A Unique Fluorescent Cocktail of Polycyclic Aromatic Hydrocarbons," *Nano Letters*, vol. 15, no. 9, pp. 6030–6035, 2015.

[144] W. Wang, B. Wang, H. Embrechts, C. Damm, A. Cadranel, V. Strauss, M. Distaso, V. Hinterberger, D. M. Guldi, and W. Peukert, "Shedding light on the effective fluorophore structure of high fluorescence quantum yield carbon nanodots," *RSC Advances*, vol. 7, no. 40, pp. 24 771–24 780, 2017.

[145] A. Cappai, C. Melis, L. Stagi, P. C. Ricci, F. Mocci, and C. M. Carbonaro, "Insight into the Molecular Model in Carbon Dots through Experimental and Theoretical Analysis of Citrazinic Acid in Aqueous Solution," *Journal of Physical Chemistry C*, vol. 125, no. 8, SI, pp. 4836–4845, 2021.

[146] W. Kasprzyk, T. Swiergosz, S. Bednarz, K. Walas, N. V. Bashmakova, and D. Bogdal, "Luminescence phenomena of carbon dots derived from citric acid and urea - a molecular insight," *Nanoscale*, vol. 10, no. 29, pp. 13 889–13 894, 2018.

[147] S. Fasbender, S. Allani, C. Wimmenauer, *et al.*, "Uptake dynamics of graphene quatum dots into primary human blood cells following in vitro exposure," *RSC Advances*, vol. 7, no. 20, pp. 12 208–12 216, 2017.

[148] S. Fasbender, L. Zimmermann, R.-P. Cadeddu, M. Luysberg, B. Moll, C. Janiak, T. Heinzel, and R. Haas, "The Low Toxicity of Graphene Quantum Dots is Reflected by Marginal Gene Expression Changes of Primary Human Hematopoietic Stem Cells," *Scientific Reports*, vol. 9, 2019.

[149] D. Kersting, S. Fasbender, R. Pilch, *et al.*, "From in vitro to ex vivo: Subcellular localization and uptake of graphene quatum dots into solid tumors," *Nanotechnology*, vol. 30, no. 39, p. 395 101, 2019.

[150] A. Kalluri, D. Debnath, B. Dharmadhikari, and P. Patra: "Chapter Twelve - Graphene Quantum Dots: Synthesis and Applications," ser. Methods in Enzymology, C. V. Kumar, Ed., vol. 609, Academic Press, 2018, pp. 335–354.

[151] P. Miao, K. Han, Y. Tang, B. Wang, T. Lin, and W. Cheng, "Recent advances in carbon nanodots: Synthesis, properties and biomedical applications," *Nanoscale*, vol. 7, no. 5, pp. 1586–1595, 2015.

[152] H. Ehtesabi, Z. Hallaji, S. Najafi Nobar, and Z. Bagheri, "Carbon dots with ph-responsive fluorescence: A review on synthesis and cell biological applications," *Microchimica Acta*, vol. 187, no. 2, 2020.

[153] P. Miao, L. Liu, Y. Li, and G. Li, "A novel electrochemical method to detect mercury (ii) ions," *Electrochemistry Communications*, vol. 11, no. 10, pp. 1904–1907, 2009.

[154] S. N. A. M. Yazid, S. F. Chin, S. C. Pang, and S. M. Ng, "Detection of Sn(II) ions via quenching of the fluorescence of carbon nanodots," *Microchimica Acta*, vol. 180, no. 1-2, pp. 137–143, 2013.

[155] T. H. Le, H. J. Lee, J. H. Kim, and S. J. Park, "Detection of Ferric Ions and Catecholamine Neurotransmitters via Highly Fluorescent Heteroatom Co-Doped Carbon Dots," *Sensors*, vol. 20, no. 12, 2020.

[156] Y. Kato, S. Ozawa, C. Miyamoto, Y. Maehata, A. Suzuki, T. Maeda, and Y. Baba, "Acidic extracellular microenvironment and cancer," *Cancer Cell International*, vol. 13, 2013.

[157] E. Boedtkjer and S. F. Pedersen: "The Acidic Tumor Microenvironment as a Driver of Cancer," *Annual Review of Physiology*, ser. Annual Review of Physiology, M. Nelson and K. Walsh, Eds., vol. 82, 2020, pp. 103–126.

[158] R. Knoblauch and C. D. Geddes, "Carbon Nanodots in Photodynamic Antimicrobial Therapy: A Review," *Materials*, vol. 13, no. 18, 2020.

[159] X. Dong, W. Liang, M. J. Meziani, Y.-P. Sun, and L. Yang, "Carbon Dots as Potent Antimicrobial Agents," *Theranostics*, vol. 10, no. 2, pp. 671–686, 2020.

[160] F. Cieplik, D. Deng, W. Crielaard, W. Buchalla, E. Hellwig, A. Al-Ahmad, and T. Maisch, "Antimicrobial photodynamic therapy - what we know and what we don't," *Critical Reviews in Microbiology*, vol. 44, no. 5, pp. 571–589, 2018.

[161] R. Klinke, H.-C. Pape, A. Kurtz, and S. Silbernagl, Eds.: *7 Blut: Ein flüssiges Organsystem*, 8. Auflage. Georg Thieme Verlag, 2018.

[162] S. M. van Neerven and L. Vermeulen, "Cell competition in development, homeostasis and cancer," *Nature Reviews Molecular Cell Biology*, vol. 24, no. 3, pp. 221–236, 2023.

[163] E. Laurenti and B. Goettgens, "From haematopoietic stem cells to complex differentiation landscapes," *Nature*, vol. 553, no. 7689, pp. 418–426, 2018.

[164] Y. K. Bozhilov, I. Hsu, E. J. Brown, and A. C. Wilkinson, "In Vitro Human Haematopoietic Stem Cell Expansion and Differentiation," *Cells*, vol. 12, no. 6, 2023.

[165] H. Cheng, Z. Zheng, and T. Cheng, "New paradigms on hematopoietic stem cell differentiation," *Protein & Cell*, vol. 11, no. 1, SI, pp. 34–44, 2020.

[166] F. Notta, S. Zandi, N. Takayama, *et al.*, "Distinct routes of lineage development reshape the human blood hierarchy across ontogeny," *Science*, vol. 351, no. 6269, p. 139, 2016.

[167] L. Velten, S. F. Haas, S. Raffel, *et al.*, "Human haematopoietic stem cell lineage commitment is a continuous process," *Nature Cell Biology*, vol. 19, no. 4, pp. 271–281, 2017.

[168] D. Karamitros, B. Stoilova, Z. Aboukhalil, *et al.*, "Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells," *Nature Immunology*, vol. 19, no. 1, pp. 85–97, 2018.

[169] J. Behrends, J. Bischofberger, R. Deutzmann, *et al.*, Eds.: *Duale Reihe - Physiologie*, 3. Auflage. Georg Thieme Verlag KG, 2016.

[170] D. Harmening: *Clinical Hematology and Fundamentals of Hemostasis*. F.A. Davis, 2024.

[171] K. Arastéh, H. Baenkler, C. Bieber, *et al.*: *Duale Reihe - Innere Medizin*, 4. Auflage. Thieme, 2018.

[172] B. Bröker, C. Schütt, and B. Fleischer: *Grundwissen Immunologie*, 4. Auflage. Springer Spektrum, 2019.

[173] S. H. E. Kaufmann: *Basiswissen Immunologie*, 1st ed., ser. Springer-Lehrbuch. Springer Berlin, Heidelberg, 2014.

[174] K. Murphy, C. Weaver, and L. Seidler: *Janeway Immunologie*. Springer Berlin Heidelberg, 2018.

[175] M. Schaich: "8 - akute leukämien," *Facharztwissen Hamatologie Onkologie (Sixth Edition)*, J. Eucker, C. W. Scholz, M. Bathon, *et al.*, Eds., Sixth Edition, Munich: Elsevier, 2024, pp. 403–430.

[176] J. D. Khoury, E. Solary, O. Abla, *et al.*, "The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms," *Leukemia*, vol. 36, no. 7, pp. 1703–1719, 2022.

[177] K. F. de Jonge-Peeters SD, de Vries EG, and V. E, "ABC transporter expression in hematopoietic stem cells and the role in AML drug resistance," *Oncology Hematology*, vol. 62, no. 3, pp. 214–226, 2007.

[178] M. Tallman, G. D.G., and J. Rowe, "Drug therapy for acute myeloid leukemia," *Blood*, vol. 106, no. 4, pp. 1154–1163, 2005.

[179] R. Robey, K. Pluchino, M. Hall, A. Fojo, S. Bates, and M. Gottesman, "Revisiting the role of efflux pumps in multidrug-resistant cancer," *Nature Reviews Cancer*, vol. 18, no. 7, pp. 452–464, 2018.

[180] D. A. Arber, A. Orazi, R. P. Hasserjian, *et al.*, "International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data," *Blood*, vol. 140, no. 11, pp. 1200–1228, 2022.

[181] J. M. Bennett, D. Catovsky, M. T. Danie, G. Flandrin, D. A. G. Galton, H. R. Gralnick, and C.Sultan, "Proposals for the Classification of the Acute Leukaemias," *British Journal of Haematology*, vol. 33, no. 4, p. 451, 1976.

[182] R. P. Hasserjian, U. Germing, and L. Malcovati, "Diagnosis and classification of myelodysplastic syndromes," *Blood*, vol. 142, no. 26, pp. 2247–2257, 2023.

[183] M. G. Della Porta, H. Tuechler, L. Malcovati, *et al.*, "Validation of WHO classification-based Prognostic Scoring System (WPSS) for myelodysplastic syndromes and comparison with the revised International Prognostic Scoring System (IPSS-R). A study of the International Working Group for Prognosis in Myelodysplasia (IWG-PM)," *Leukemia*, vol. 29, no. 7, pp. 1502–1513, 2015.

[184] L. D. Yuen and R. P. Hasserjian, "Morphologic Characteristics of Myelodysplastic Syndromes," *Clinics in Laboratory Medicine*, vol. 43, no. 4, pp. 577–596, 2023.

[185] A. S. Kubasch and U. Platzbecker: "9 - Myelodysplastische Neoplasien (MDS)," *Facharztwissen Hamatologie Onkologie (Sixth Edition)*, J. Eucker, C. W. Scholz, M. Bathon, *et al.*, Eds., Sixth Edition, Munich: Elsevier, 2024, pp. 431–446.

[186] J. J. Rennick, A. P. R. Johnston, and R. G. Parton, "Key principles and methods for studying the endocytosis of biological and nanoparticle therapeutics," *Nature Nanotechnology*, vol. 16, no. 3, pp. 266–276, 2021.

[187] M. S. de Almeida, E. Susnik, B. Drasler, P. Taladriz-Blanco, A. Petri-Fink, and B. Rothen-Rutishauser, "Understanding nanoparticle endocytosis to improve targeting strategies in nanomedicine," *Chemical Society Reviews*, vol. 50, no. 9, pp. 5397–5434, 2021.

[188] S. Zhang, H. Gao, and G. Bao, "Physical principles of nanoparticle cellular endocytosis," *ACS Nano*, vol. 9, no. 9, pp. 8655–8671, 2015.

[189] X. P. Lin, J. D. Mintern, and P. A. Gleeson, "Macropinocytosis in Different Cell Types: Similarities and Differences," *Membranes*, vol. 10, no. 8, 2020.

[190] M. Kaksonen and A. Roux, "Mechanisms of clathrin-mediated endocytosis," *Nature Reviews Molecular Cell Biology*, vol. 19, no. 5, pp. 313–326, 2018.

[191] J. J. Thottacherry, M. Sathe, C. Prabhakara, and S. Mayor: "Spoiled for choice: Diverse endocytic pathways function at the cell surface," R. Lehmann, Ed.

[192] G. J. Doherty and H. T. McMahon, "Mechanisms of Endocytosis," *Annual Review of Biochemistry*, vol. 78, pp. 857–902, 2009.

[193] J. Zhao and M. H. Stenzel, "Entry of nanoparticles into cells: The importance of nanoparticle properties," *Polymer Chemistry*, vol. 9, no. 3, pp. 259–272, 2018.

[194] M. Ehrlich, W. Boll, A. van Oijen, R. Hariharan, K. Chandran, M. Nibert, and T. Kirchhausen, "Endocytosis by random initiation and stabilization of clathrin-coated pits," *Cell*, vol. 118, no. 5, pp. 591–605, 2004.

[195] H. T. McMahon and E. Boucrot, "Molecular mechanism and physiological functions of clathrin-mediated endocytosis," *Nature Reviews Molecular Cell Biology*, vol. 12, no. 8, pp. 517–533, 2011.

[196] E. Boucrot, A. P. A. Ferreira, L. Almeida-Souza, S. Debard, Y. Vallis, G. Howard, L. Bertot, N. Sauvonnet, and H. T. McMahon, "Endophilin marks and controls a clathrin-independent endocytic pathway," *Nature*, vol. 517, no. 7535, pp. 460–465, 2015.

[197] A. P. A. Ferreira, A. Casamento, S. C. Roas, *et al.*, "Cdk5 and gsk3ß inhibit fast endophilin-mediated endocytosis," *Nature Communications*, vol. 12, no. 1, 2021.

[198] A. Casamento and E. Boucrot, "Molecular mechanism of Fast Endophilin-Mediated Endocytosis," *Biochemical Journal*, vol. 477, no. 12, pp. 2327–2345, 2020.

[199] M. Kirkham, A. Fujita, R. Chadda, S. Nixon, T. Kurzchalia, D. Sharma, R. Pagano, J. Hancock, S. Mayor, and R. Parton, "Ultrastructural identification of uncoated caveolin-independent early endocytic vehicles," *Journal of Cell Biology*, vol. 168, no. 3, pp. 465–476, 2005.

[200] P. Moreno-Layseca, N. Z. Jantti, R. Godbole, *et al.*, "Cargo-specific recruitment in clathrin- and dynamin-independent endocytosis," *Nature Cell Biology*, vol. 23, no. 10, pp. 1073–1084, 2021.

[201] E. Uribe-Querol and C. Rosales, "Phagocytosis: Our Current Understanding of a Universal Biological Process," *Frontiers in Immunology*, vol. 11, 2020.

[202] J. Huotari and A. Helenius, "Endosome maturation," *EMBO Journal*, vol. 30, no. 17, pp. 3481–3500, 2011.

[203] R. Vacha, F. J. Martinez-Veracoechea, and D. Frenkel, "Receptor-Mediated Endocytosis of Nanoparticles of Various Shapes," *Nano Letters*, vol. 11, no. 12, pp. 5391–5395, 2011.

[204] B. Chithrani, A. Ghazani, and W. Chan, "Determining the size and shape dependence of gold nanoparticle uptake into mammalian cells," *Nano Letters*, vol. 6, no. 4, pp. 662–668, 2006.

[205] A. C. Anselmo, M. Zhang, S. Kumar, D. R. Vogus, S. Menegatti, M. E. Helgeson, and S. Mitragotri, "Elasticity of Nanopartides Influences Their Blood Circulation, Phagocytosis, Endocytosis, and Targeting," *ACS Nano*, vol. 9, no. 3, pp. 3169–3177, 2015.

[206] X. Yi and H. Gao, "Kinetics of receptor-mediated endocytosis of elastic nanoparticles," *Nanoscale*, vol. 9, no. 1, pp. 454–463, 2017.

[207] P. Gurnani, C. Sanchez-Cano, H. Xandri-Monje, J. Zhang, S. H. Ellacott, E. D. H. Mansfield, M. Hartlieb, R. Dallmann, and S. Perrier, "Probing the Effect of Rigidity on the Cellular Uptake of Core-Shell Nanoparticles: Stiffness Effects are Size Dependent," *Small*, vol. 18, no. 38, 2022.

[208] P. Foroozandeh and A. A. Aziz, "Insight into Cellular Uptake and Intracellular Trafficking of Nanoparticles," *Nanoscale Research Letters*, vol. 13, 2018.

[209] N. Means, C. K. Elechalawar, W. R. Chen, R. Bhattacharya, and P. Mukherjee, "Revealing macropinocytosis using nanoparticles," *Molecular Aspects of Medicine*, vol. 83, no. SI, 2022.

[210] C. Goodman, C. McCusker, T. Yilmaz, and V. Rotello, "Toxicity of gold nanoparticles functionalized with cationic and anionic side chains," *Bioconjugate Chemistry*, vol. 15, no. 4, pp. 897–900, 2004.

[211] J. Lovric, H. Bazzi, Y. Cuie, G. Fortin, F. Winnik, and D. Maysinger, "Differences in subcellular distribution and toxicity of green and red emitting CdTe quantum dots," *Journal of Molecular Medicine*, vol. 83, no. 5, pp. 377–385, 2005.

[212] C. He, Y. Hu, L. Yin, C. Tang, and C. Yin, "Effects of particle size and surface charge on cellular uptake and biodistribution of polymeric nanoparticles," *Biomaterials*, vol. 31, no. 13, pp. 3657–3666, 2010.

[213] N. D. Donahue, H. Acar, and S. Wilhelm, "Concepts of nanoparticle cellular uptake, intracellular trafficking, and kinetics in nanomedicine," *Advanced Drug Delivery Reviews*, vol. 143, no. SI, pp. 68–96, 2019.

[214] H. Appelqvist, P. Waster, K. Kagedal, and K. Ollinger, "The lysosome: From waste bag to potential therapeutic target," *Journal of Molecular Cell Biology*, vol. 5, no. 4, pp. 214–226, 2013.

[215] J. P. Luzio, P. R. Pryor, and N. A. Bright, "Lysosomes: Fusion and function," *Nature Reviews Molecular Cell Biology*, vol. 8, no. 8, pp. 622–632, 2007.

[216] A. Ballabio and J. S. Bonifacino, "Lysosomes as dynamic regulators of cell and organismal homeostasis," *Nature Reviews Molecular Cell Biology*, vol. 21, no. 2, pp. 101–118, 2020.

[217] C. Settembre, A. Fraldi, D. L. Medina, and A. Ballabio, "Signals from the lysosome: A control centre for cellular clearance and energy metabolism," *Nature Reviews Molecular Cell Biology*, vol. 14, no. 5, pp. 283–296, 2013.

[218] H. Xu and D. Ren: "Lysosomal Physiology," *Annual Review of Physiology*, ser. Annual Review of Physiology, D. Julius, Ed., vol. 77, 2015, pp. 57–80.

[219] F. Wang, R. Gomez-Sintes, and P. Boya, "Lysosomal membrane permeabilization and cell death," *Traffic*, vol. 19, no. 12, pp. 918–931, 2018.

[220] R. M. Perera and R. Zoncu: "The Lysosome as a Regulatory Hub," *Annual Review of Cell and Developmental Biology*, ser. Annual Review of Cell and Developmental Biology, R. Schekman, Ed., vol. 32, 2016, pp. 223–253.

[221] A. Ballabio and V. Gieselmann, "Lysosomal disorders: From storage to cellular damage," *Biochimica et Biophysica Acta: Molecular Cell Research*, vol. 1793, no. 4, SI, pp. 684–696, 2009.

[222] F. M. Platt, "Emptying the stores: Lysosomal diseases and therapeutic strategies," *Nature Reviews Drug Discovery*, vol. 17, no. 2, pp. 133–150, 2018.

[223] S. R. Bonam, F. Wang, and S. Muller, "Lysosomes as a therapeutic target," *Nature Reviews Drug Discovery*, vol. 18, no. 12, pp. 923–948, 2019.

[224] B. Levine and G. Kroemer, "Biological Functions of Autophagy Genes: A Disease Perspective," *Cell*, vol. 176, no. 1-2, pp. 11–42, 2019.

[225] I. Dikic and Z. Elazar, "Mechanism and medical implications of mammalian autophagy," *Nature Reviews Molecular Cell Biology*, vol. 19, no. 6, pp. 349–364, 2018.

[226] A. Stolz, A. Ernst, and I. Dikic, "Cargo recognition and trafficking in selective autophagy," *Nature Cell Biology*, vol. 16, no. 6, pp. 495–501, 2014.

[227] L. Galluzzi, J. M. Bravo-San Pedro, B. Levine, D. R. Green, and G. Kroemer, "Pharmacological modulation of autophagy: Therapeutic potential and persisting obstacles," *Nature Reviews Drug Discovery*, vol. 16, no. 7, pp. 487–511, 2017.

[228] S. R. Yoshii and N. Mizushima, "Monitoring and Measuring Autophagy," *International Journal of Molecular Sciences*, vol. 18, no. 9, 2017.

[229] D. Gatica, V. Lahiri, and D. J. Klionsky, "Cargo recognition and degradation by selective autophagy," *Nature Cell Biology*, vol. 20, no. 3, pp. 233–242, 2018.

[230] V. Rogov, V. Doetsch, T. Johansen, and V. Kirkin, "Interactions between Autophagy Receptors and Ubiquitin-like Proteins Form the Molecular Basis for Selective Autophagy," *Molecular Cell*, vol. 53, no. 2, pp. 167–178, 2014.

[231] Y. Katsuragi, Y. Ichimura, and M. Komatsu, "p62/SQSTM1 functions as a signaling hub and an autophagy adaptor," *FEBS Journal*, vol. 282, no. 24, pp. 4672–4678, 2015.

[232] M. J. Wilkerson, "Principles and Applications of Flow Cytometry and Cell Sorting in Companion Animal Medicine," *Veterinary Clinics of North America: Small Animal Practice*, vol. 42, no. 1, pp. 53–71, 2012.

[233] B. Cormack, R. Valdivia, and S. Falkow, "FACS-optimized mutants of the green fluorescent protein (GFP)," *Gene*, vol. 173, no. 1, pp. 33–38, 1996.

[234] P. S. Sonal M Manohar and A. Nair, "Flow Cytometry: Principles, Applications and Recent Advances," *Bioanalysis*, vol. 13, no. 3, pp. 181–198, 2021.

[235] P. Dey: *Diagnostic Flow Cytometry in Cytology.* Springer Singapore, 2021.

[236] C. E. Pedreira, E. S. Costa, Q. Lecrevisse, J. J. M. van Dongen, A. Orfao, and E. Consortium, "Overview of clinical flow cytometry data analysis: Recent advances and future challenges," *Trends in Biotechnology*, vol. 31, no. 7, pp. 415–425, 2013.

[237] J. Brummelman, C. Haftmann, N. G. Núñez, G. Alvisi, E. M. C. Mazza, B. Becher, and E. Lugli, "Development, application and computational analysis of high-dimensional fluorescent antibody panels for single-cell flow cytometry," *Nature Protocols*, vol. 14, no. 7, pp. 1946–1969, 2019.

[238] D. Qu, M. Zheng, J. Li, Z. Xie, and Z. Sun, "Tailoring color emissions from N-doped graphene quatum dots for bioimaging applications," *Light: Science & Applications*, vol. 4, no. 12, e364–e364, 2015.

[239] M. Sudolska, M. Dubecky, S. Sarkar, C. J. Reckmeier, R. Zboril, A. L. Rogach, and M. Otyepka, "Nature of Absorption Bands in Oxygen-Functionalized Graphitic Carbon Dots," *Journal of Physical Chemistry C*, vol. 119, no. 23, pp. 13 369–13 373, 2015.

[240] S. Fasbender: "The interaction of graphene quantum dots with human cells," Ph.D. dissertation, Heinrich-Heine-Universität, 2020.

[241] G. Rajender and P. K. Giri, "Formation mechanism of graphene quantum dots and their edge state conversion probed by photoluminescence and Raman spectroscopy," *Journal of Materials Chemistry C*, vol. 4, no. 46, pp. 10 852–10 865, 2016.

[242] A. Williams, S. Winfield, and J. N. Miller, "Relative Fluorescence Quantum Yields Using a Computer-controlled Luminescence Spectrometer," *Analyst*, vol. 108, no. 1290, pp. 1067–1071, 1983.

[243] T. Tkaczyk: *Field Guide to Microscopy*, ser. Field Guides. SPIE Press, 2010.

[244] D. Boas, C. Pitris, and N. Ramanujam, Eds.: *Handbook of Biomedical Optics*. CRC Press, 2012.

[245] J. Jonkman, C. M. Brown, G. D. Wright, K. I. Anderson, and A. J. North, "Tutorial: Guidance for quantitative confocal microscopy," *Nature Protocols*, vol. 15, no. 5, pp. 1585–1611, 2020.

[246] J. G. H. Towbin T. Staehelin, "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, no. 9, pp. 4350–4354, 1979.

[247] B. T. Kurien: *Western Blotting for the Non-Expert*, A. E. Kalyuzhny, Ed., ser. Techniques in Life Science and Biomedicine for the Non-Expert. Springer Cham, 2021.

[248] W. Luttmann, K. Bratke, M. Küpper, and D. Myrtek: *Der Experimentator: Immunologie*, 4., vollständig überarbeitete und korrigierte Auflage, ser. Experimentator. Springer Spektrum Berlin, Heidelberg, 2014.

[249] B. T. Kurien and R. H. Scofield, "Western blotting," *Methods*, vol. 38, no. 4, pp. 283–293, 2006.

[250] P. R. Reddy and N. Raju: "Gel-Electrophoresis and Its Applications," *Gel Electrophoresis*, S. Magdeldin, Ed., Rijeka: IntechOpen, 2012, ch. 2.

[251] J. H. Gross: *Massenspektrometrie*. Springer Spektrum Berlin, Heidelberg, 2019, Spektroskopiekurs kompakt.

[252] P. Christen, R. Jaussi, and R. Benoit: "Trennverfahren und allgemeine analysemethoden," *Biochemie und Molekularbiologie: Eine Einführung in 40 Lerneinheiten*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2024.

Abbreviations

AFM	Atomic Force Microscopy
AI	Artificial Intelligence
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
APDs	Avalanche Photodiodes
AUC	Area Under The Curve
BM	Bone Marrow
bPEI	Branched Polyethylenimine
CD	Cluster Of Differentiation
CCD	Charge-Coupled Device
CLIC/GEEC	Clathrin-Independent/Dynamin-Independent Endocytosis
CLL	Chronic Lymphocytic Leukemia
CLPs	Common Lymphoid Progenitor Cells
CME	Clathrin-Mediated Endocytosis
CMPs	Common Myeloid Progenitor Cells
CNDs	Carbon Nanodots
CNN	Convolution Neural Network
CQDs	Carbon Quantum Dots
DETA	Diethylenetriamine
DL	Deep Learning
DNNs	Deep Neural Networks
DOX	Doxorubicin
DR	Dimensionality Reduction
EDC	1-Ethyl-3-(3-Dimethylaminopropyl)Carbodiimide
FACS	Fluorescence-Activated Cell Sorter
FEME	Fast Endophilin-Mediated Endocytosis
FE	Feature Extraction
FEAs	Feature Extracting Algorithms
FlowCAP	Flow Cytometry: Critical Assessment Of Population Identification Methods
FlowSOM	Flow-Self-Organizing-Maps
FRET	Förster Resonance Energy Transfer
FS	Feature Selection

FSC	Forward-Scattered Light
GMPs	Granulocyte-Macrophage Progenitors
GQDs	Graphene Quantum Dots
HSCs	Hematopoietic Stem Cells
HSPCs	Hematopoietic Stem And Progenitor Cells
ICC	International Consensus Classification
ISOMAP	Isometric Mapping
KL	Kullback-Leibler
LC3	Microtubule-Associated Protein 1 Light Chain 3B
LDA	Linear Discriminant Analysis
LMPPs	Lymphoid-Primed Multipotential Progenitor Cells
LSCs	Leukemic Stem Cells
LSDs	Lysosomal Storage Disorders
LT-HSCs	Long-Term HSCs
MDS	Myelodysplastic Neoplasms
MEPs	Megakaryocyte-Erythrocyte Progenitors
ML	Machine Learning
MPPs	Multipotent Progenitor Cells
MRD	Minimal Residual Disease
MS	Mass Spectrometry
NHS	N-Hydroxysuccinimide
NK	Natural Killer
PAH	Polycyclic Aromatic Hydrocarbon
PCA	Principal Component Analysis
PBS	Phosphate-Buffered Saline
PCs	Principal Components
PDI	Photodynamic Inactivation
PMTs	Photo-Multiplier Tubes
ROS	Reactive Oxygen Species
SDS	Sodium Dodecyl Sulfate
SNE	Stochastic Neighbor Embedding
SSC	Side-Scattered Light
ST-HSCs	Short-Term HSCs
TBM	Tight-Binding Model
TEM	Transmission Electron Microscopy
t-SNE	t-Distributed Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation And Projection
XPS	X-Ray Photoelectron Spectroscopy

Danksagung

dir vielen Dank für deine Hilfsbereitschaft und die schnelle Unterstützung bei Fragen.

Vielen Dank an unsere Kooperationspartner Prof. Dr. Laura Hartmann, Dr. Stephen Hill und Dr. Serap Üçlü aus der Arbeitsgruppe für Makromolekulare Chemie für die gemeinsame Arbeit an den Zucker-funktionalisierten Nanopartikeln.

Bedanken möchte ich mich auch bei dem Molecular Proteomics Laboratory des Biologisch-Medizinischen Forschungszentrum an der HHU unter der Leitung von Prof. Dr. Kai Stühler für die Durchführung der Proteom- und Sekretomanalysen. Insbesondere möchte ich mich dafür bei Thomas Lenz für die Durchführung und die umfangreiche Expertise und das Engagement bei der Auswertung bedanken.

Ein besonderen Dank möchte ich der Jürgen Manchot Stiftung aussprechen für die langjährige Förderung während meiner Promotion, die sehr hilfreich war und ohne die das ein oder andere Experiment in dem Umfang nicht möglich gewesen wäre.

Ich möchte mich auch bei dem Team des Center of Advanced Imaging (CAi) der HHU bedanken, für die Bereitstellung der Mikroskope für unsere Experimente und die Einführungen in diese Geräte.

Bei meinem Studienfreund, späteren Betreuer und zum Schluss Bürokollegen Dr. Christian Wimmenauer möchte ich mich auch ganz herzlich bedanken. Christian, ein ganz großes Dankeschön für deine Unterstützung und Hilfe seit dem gemeinsamen Studienbeginn, die zahlreichen gemeinsamen Stunden im Hörsaal, Labor und Büro, es war immer schön. Danke, dass ich von deinem umfangreichen Wissen profitieren und lernen durfte. Wenn wir nicht auf dem Weihnachtsmarkt über eine mögliche Masterarbeit gesprochen hätten, wäre ich wahrscheinlich nicht ans Institut gekommen und wenn du damals nicht dein Interesse an AI in die Arbeit hättest einfließen lassen, wäre auch diese Doktorarbeit so nie zustande gekommen, ich danke dir.

Ebenso möchte ich mich bei meinem Vorgänger und Betreuer meiner Masterarbeit Dr. Stefan Fasbender herzlich bedanken. Danke Stefan, dass du mich in die Arbeit am Durchflusszytometer und im Labor eingeführt hast und für deine Vorarbeiten zu den CNDs, auf denen ich aufbauen durfte. Danke für deine Unterstützung und auch die ein oder anderer Erinnerung, Fünfe gerade sein zu lassen.

Ein großer Dank geht auch an alle Studierenden, mit denen ich im Rahmen ihrer Abschlussarbeiten im Laufe der Jahre zusammenarbeiten durfte: Faranak, Daniela, Karin, Rika, Maxime, Carla, Frederik, Simon, Sarah und Carina. Danke für die tolle Zusammenarbeit, es hat mir viel Freude bereitet. Ebenso möchte ich mich bei den zahlreichen Studierenden bedanken, die ich im Rahmen von Praktika und Übungsgruppen begleiteten durfte, ich hatte immer viel Spaß dabei.

Bianka Lindenau möchte ich danken für die Übernahme der vielen administrativen Aufgaben. Vielen Dank Bianka, dass du alles bestmöglich vorbereitet und organisiert hast und ich meistens nur noch unterschreiben musste und danke für deine ansteckende gute Laune.

Natürlich dürfen meine (ehemaligen) Doktoranden- und Postdoc-Kolleg:innen vom Lehrstuhl für Festkörperphysik nicht fehlen. Stefan, Laurin, Svenja, Christian, Lukas, Carla,

Frederik und Mihai, ich danke euch für die tolle Atmosphäre am Lehrstuhl, den Kuchen, den Austausch, die Unterstützung und die spaßigen Stunden auch außerhalb der Arbeit. Ohne euch wären die letzten Jahre sicher nicht so schön gewesen, wie sie waren.

Carla, Christian und Markus möchte ich auch ganz herzlich danken für das Korrekturlesen der Arbeit und die vielen hilfreichen Anmerkungen.

Zuletzt möchte ich mich bei meiner Familie und meinen Freund:innen bedanken, ohne die das alles gar nicht möglich gewesen wäre. Meinen Eltern Ulrike und Markus möchte ich für ihre Liebe und Unterstützung danken, insbesondere während der sehr herausfordernden letzten Jahre. Danke, dass ich mich immer auf euch verlassen kann und ihr immer für mich da seid, insbesondere du, Mama! Meinen Freund:innen möchte ich danken, dass ihr mich - teilweise schon fast mein ganzes Leben lang - begleitet und ich mich immer auf euch verlassen kann.