# Supporting Information

"Simultaneous cathodoluminescence and electron microscopy cytometry of cellular vesicles labeled with fluorescent nanodiamonds"

S. Nagarajan *et al.*

**Contamination of Type 2 ND sample by particles containing red emitting NV centers.**

In green light emitting type 2 sample we have observed the presence of red light emitting nanoparticles in addition to the green emitting ones. To determine the contamination of type 2 sample, we dispersed ND on a carbon film and imaged them by CL-STEM. Figure S1 shows that type 2 sample diamond nanocrystals displaying a CL signal contain either green (H3) or red (NV) color centers but never both in the same crystal. From the analysis of CL spectral images for 8 field of view of size 15.17 µm x 15.17 µm we inferred that 57 of type 2 ND are purely green emitters.

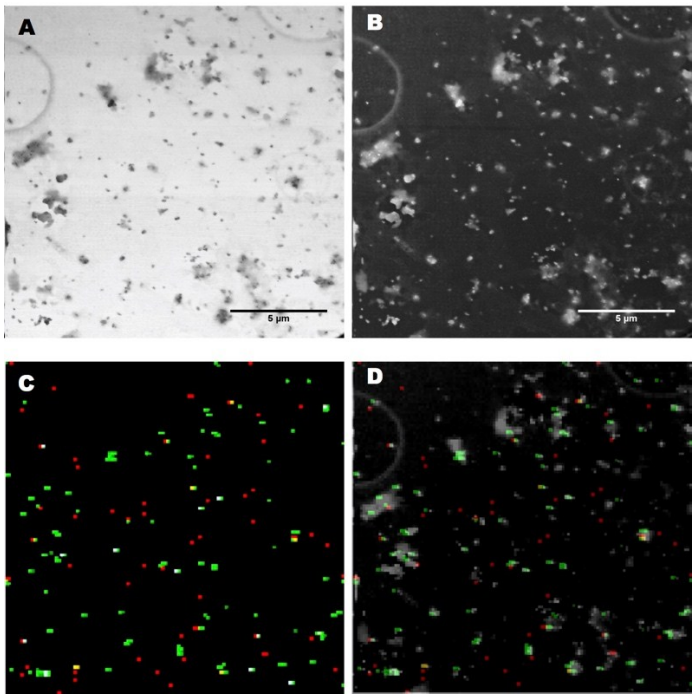


**Figure S1 – Contamination of Type 2 ND sample by particles containing red emitting NV centers.** (A) Low magnification bright field image of Type 2 NDs. (B), corresponding dark field image for the same region with a smaller sampling (C), corresponding CL emission. (D) Overlap of B and C demonstrating the presence of contaminating red emitters in Type 2 sample. Scale bar: 5 µm.

As stated in the main text, the Type 2 ND sample is unfortunately not spectrally pure and contains either red or green emitting nanoparticles. The red emitters are NV centers in type 1 and in type 2 NDs, and we did not observe significant CL spectrum differences in **bare** ND of both types. However, we noticed that **after cationic polymer coating**, the spectrum of embedded NV centers were modified differently, depending on the type of polymer used (Figure 2C of the main text). The phonon region is indeed significantly different in both types of NDs.

In order to distinguish between PAH-coated Type 1 NDs (which are all red emitters) and PEI-coated Type 2 red-emitting NDs, we have developed a method relying on **multivariate analysis**. Specifically, we have applied K-means analysis to the coloaded sample. This analysis has lead to the discrimination of the NDs into two clusters. These clusters could be assigned to either ND-PAH or ND-PEI by comparison with the results of the same analysis done on reference samples made up of only ND-PAH or ND-PEI with *a priori* knowledge of polymer coating.

**K-Means method used to identify the coating of red emitting NDs.**

Principle – K-means is a clustering method that groups different data into separate clusters, based on the minimal distance between observations within a given cluster. The number of clusters is an input of the method, and thus an evaluation of the relevance of the choice of a particular number of clusters is necessary. We note that several numbers of clusters may be optimal, as sub-clustering might arise in samples. Moreover, the K-Means relying on heuristic minimizing methods, might be prone to initial condition errors, and it is therefore usually run several times with different initial conditions to avoid local minima.

In the present work, the data set is made of the CL spectra, and each cluster is centred around a centroid spectrum, from which any spectrum comprised into the given cluster is supposed to be the closest, and any spectrum comprised in any other cluster supposed to be furthest.

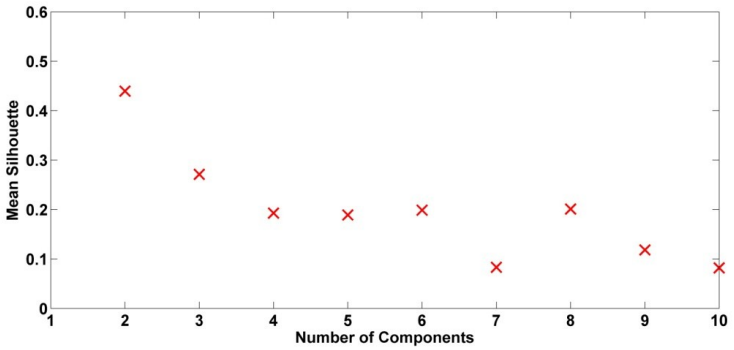


**Figure S2** – Mean Silhouette for a K-Means analysis of polycation coated NDs in the coloaded samples.

Here, the criterion that is used to validate the optimal number of clusters in the sample pool is a maximal mean silhouette index[1] (Figure S2). The **silhouette index** defines how well separated the sample groups are. A high silhouette index for a given number of groups indicates that the different spectra cluster well around their centroid and are far away from the other centroids. Therefore, the optimal number of components has the maximum mean silhouette. If this optimum is reached for different number of clusters, sub-clustering is taking place. Moreover, some reports suggest a relationship between the values of the mean silhouette and the strength of the difference between samples. Values above 0.8 indicate a well defined structure, 0.5 to 0.8 presence of some structure, 0.3-0.5 weak structure and less than 0.3 might be an artefact. But these values are empirical based on observations that strongly depend on the analysed sample and cannot be used as a rule of thumb. Finally, the physical content of the output of the automatic method must be also carefully considered with respect to relevant physical parameters.

In practice, the CL spectra were analysed using the K-Means function of MATLAB statistical analysis tool box. K-Means algorithm groups the spectra in clusters for which it calculates the mean centroid spectrum representing best each cluster (see Figure S3 for an example). A minimum of 3 replicates (starting values are changed by the method in MATLAB) were run to ensure identification of a global solution by the K-means method. The number of clusters were chosen to be two, in agreement with the silhouette behaviour (Figure S2) as a function of the number of clusters, and the known physical input (i.e. the existence of two different coatings).

Applying the method of identification of the coating to blind samples – The method was first applied to cells co-incubated with ND-PAH and ND-PEI. We used K-Means to analyse CL spectra of red emitting ND only. This led to the finding of two clusters of red spectra (each presented as 2D stacks on Figure S3, left.), with their corresponding centroids represented in the centre of Figure S3. The corresponding residuals are represented on the right side, while the difference between the two centroids is shown at the bottom of Figure S3. It is clear that both centroids differ by the ratio between the phonon region and the zero-phonon line, as can be seen in Figure 2C of the main text, but here retrieved with limited prior information or assumption. We now need a criterion to assign a given centroid, and thus cluster to each type of coating.

[1] Rousseeuw, P. J. *J. Comput. Appl. Math.* **1987**, *20*, 53–65.

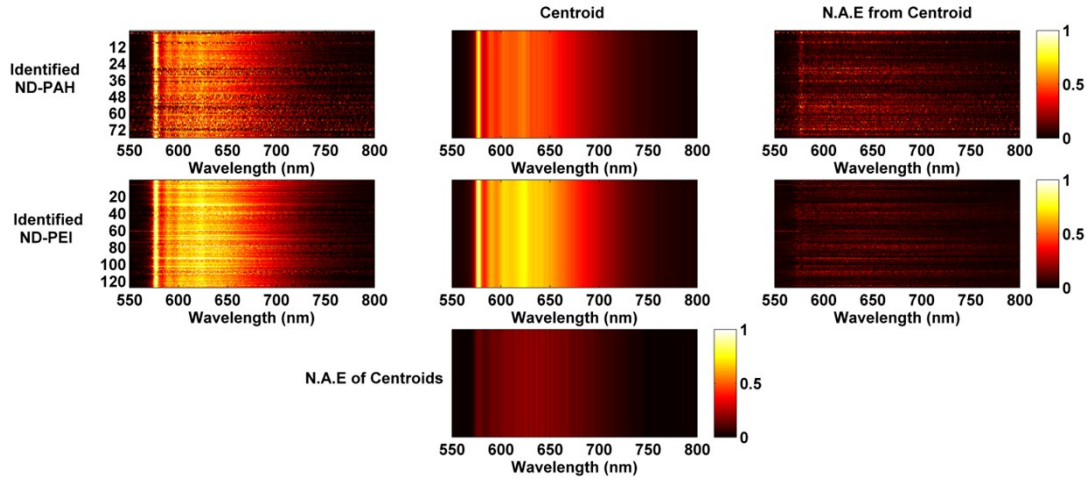


**Figure S3 –** K-Means analysis of CL spectra of red-emitting ND in the case of cell co-incubated with ND-PAH and ND-PEI . In this representation, each spectrum is represented as a linear image. All spectra are then stacked on each other to form a two dimensional image. This representation helps to identify differences and similarities between spectra. Top row : first cluster, further attributed to ND-PAH signature. From left to right, experimental spectra, centroid and NAE: Normalized absolute error (normalized value of the difference between experimental and centroid spectra). The centroid 2D image is the repetition of the centroid spectrum. Middle row: second cluster, further attributed to ND-PEI. Bottom: N.A.E between the two centroids .

Applying the methods of polymer coating identification to reference samples – We applied K-Means to a set of data gathering CL spectra of red emitting ND from known samples, one with only ND-PEI and the other with only ND-PAH (Figure S4). The purpose of this test was to (i) verify the accuracy of K-Means to identify clusters of each type of coated NDs and (ii) assign a typical centroid to a given coating. All the ND-PEI have been correctly assigned (100% success) while ≈60 % of the ND-PAH have been correctly assigned. Moreover, the ND-PEI centroid spectrum differs from the ND-PAH one by a larger phonon contribution. As a confirmation, we have compared the centroid spectra of red emitting ND from samples containing pure ND-PEI or pure ND-PAH, and we have observed the same differences (data not shown). Such an observation allowed us to assign each of the centroids of the co-incubated cell sample in Figure S3 to a given type of ND-coating.

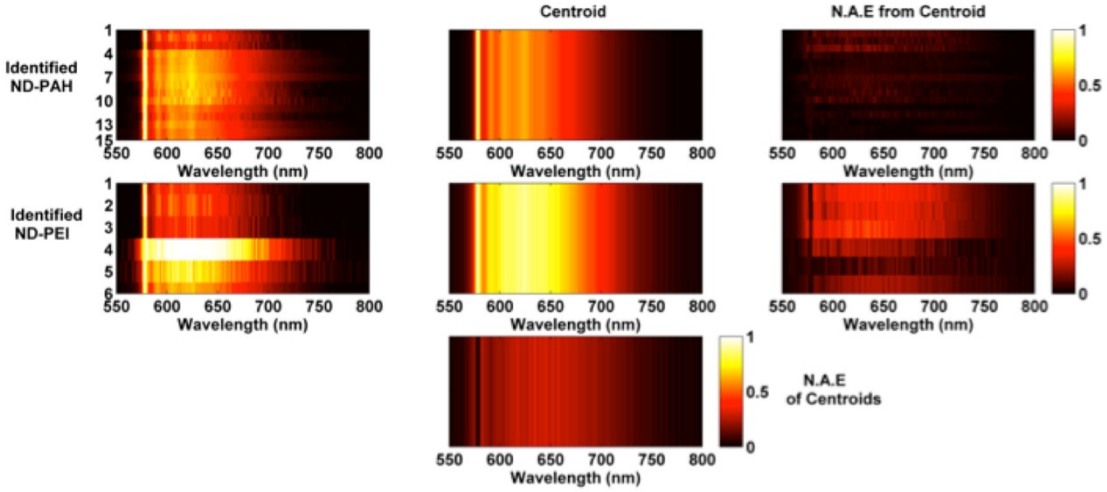


**Figure S4 –** K-Means analysis for the reference sample consisting of a data set made of CL spectra of nanodiamonds with known coating. K-Means is able to identify the two populations of coated nanodiamonds. Top row : First cluster, corresponding to the ND-PAH signature. From left to right, experimental spectra, centroid and N.A.E from centroid. The centroid 2D image is the repetition of the centroid spectrum. Middle row: second cluster, corresponding to the ND-PEI signature. Bottom: N.A.E between the two centroids.