Supplementary information for "Statistical physics approach to quantifying differences in myelinated nerve fibers"

César H. Comin,[1,*] João R. Santos,[2,†] Dario Corradini,[2,‡] Will Morrison,[2] Chester Curme,[2] Douglas L. Rosene,[3] Andrea Gabrielli,[4,5] Luciano da F. Costa,[1] and H. Eugene Stanley[2]

[1]*Institute of Physics at São Carlos, University of São Paulo, São Carlos, SP 13560-970, Brazil*

[2]*Center for Polymer Studies and Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA*

[3]*Boston University, School of Medicine, Department of Anatomy & Neurobiology, Boston, Massachusetts 02118, USA*

[4]*Istituto dei Sistemi Complessi (ISC) - CNR, UOS "Sapienza", Dipartimento di Fisica, "Sapienza" Università di Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy*

[5]*IMT Alti Studi Lucca, Piazza S. Ponziano 6, 55100 Lucca, Italy*

(Dated: January 24, 2014)

* These authors contributed equally to this work

† These authors contributed equally to this work; Correspondence to: jrsantos@bu.edu

‡ Present address: Sorbonne Universités, UPMC Univ Paris 06, UMR 8234, PHENIX, Paris, France & CNRS, UMR 8234, PHENIX, Paris, France

TABLE S1. **List of the features considered.** The table lists all the 45 features employed as input in the feature selection and the corresponding abbreviations used. For each feature a brief definition is provided where appropriate, together with the corresponding symbol/mathematical formula. The symbol $\langle \cdots \rangle$ corresponds to the average taken over the myelinated axons contained in each individual EM image.

Feature	**Abbreviation**	**Description**	**Symbol/Formula**
Number of axons	NumAxons	Number of centroids of myelinated axons inside a fixed and equal area $\mathrm{A_S}$ for all samples.	N
Density (μm^{-2})	Density	Density of myelinated axons, i.e. number of centroids of myelinated axons inside a fixed and equal area for all samples, divided by this sample area $\mathrm{A_S}$.	$\rho = \frac{N}{\mathrm{A_S}}$
Fraction of occupied area	FracOccupArea	Sum of the cross-sectional area occupied by the myelinated axons divided by the fixed sample area $\mathrm{A_S}$ used to calculate the axon density.	$f = \frac{\sum_i^N A_i}{\mathrm{A_S}}$
Mean area (μm^2)	MeanArea	Average cross-sectional area of the myelinated axons.	$\langle A \rangle = \frac{1}{N} \sum_i^N A_i$
Standard deviation of the area (μm^2)	StdDevArea	Standard deviation of the cross-sectional areas of the myelinated axons.	σ_A
Coefficient of variation of the area	CoeffVariation	Coefficient of variation of the myelinated axon areas.	$\frac{\sigma_A}{\langle A \rangle}$
Skewness of the area distribution	Skewness	Skewness of the area distribution of myelinated axons.	$\left\langle \left(\frac{A - \langle A \rangle}{\sigma_A} \right)^3 \right\rangle$
Mean perimeter (μm)	MeanPerimeter	Mean perimeter of the myelinated axons.	$\langle P \rangle$
Standard deviation of the perimeter (μm)	StdDevPerimeter	–	σ_P
Mean elongation	MeanElongation	The elongation is the ratio between the largest and smallest eigenvalue of the covariance matrix of the shape contour. For a circle, $E = 1$.	$\langle E \rangle$
Standard deviation of the elongation	StdDevElongation	–	σ_E
Mean circularity	MeanCircularity	The circularity measures how close is the shape of an axon to a circle. For a circle, $C = 1$.	$\langle C \rangle = \left\langle \frac{4\pi A}{P^2} \right\rangle$
Standard deviation of the circularity	StdDevCircularity	–	σ_C
Mean diameter (μm)	MeanDiameter	Largest distance between any two contour points.	$\langle D \rangle$
Standard deviation of the diameter	StdDevDiameter	–	σ_D

TABLE S1. **List of the features considered.** (continued)

Mean absolute curvature	MeanAbsCurvature	For each point $\mathcal{P}$ in the perimeter of an axon, the curvature $\kappa_{\mathcal{P}}$ at that point is defined as the inverse of the radius of the circle approximating the perimeter curve around that point. The mean absolute curvature of an axon is defined as the average curvature over all perimeter points of that axon times its radius: $K = \langle \kappa_{\mathcal{P}} \rangle_{\mathcal{P}} \cdot R$, where the axon radius is given by $R = \sqrt{A/\pi}$.	$\kappa_{\mathcal{P}} = \frac{\lvert \dot{x}\ddot{y} - \ddot{x}\dot{y} \rvert}{(\dot{x}^2 + \dot{y}^2)^{\frac{3}{2}}}$ $\langle K \rangle = \langle \langle \kappa_{\mathcal{P}} \rangle_{\mathcal{P}} R \rangle$
Standard deviation of the curvature	StdDevCurvature	–	σ_K
Mean bending energy	MeanBendEnergy	The bending energy of a contour is the sum of the squared curvature at each point of the countour divided by its perimeter.	$\langle B \rangle = \left\langle \frac{\sum_{\mathcal{P}} \kappa_{\mathcal{P}}^2}{P} \right\rangle$
Standard deviation of the bending energy	StdDevBendEnergy	–	σ_B
1st nearest neighbor mean distance (µm)	1stNNDist	Average distance from each myelinated axon to the closest myelinated axon.	$\langle r_1 \rangle$
Standard error of the 1st nearest neighbor mean distance (µm)	StdErr1stNNDist	Error of 1stNNDist.	Σ_{r_1}
Standard deviation of the 1st nearest neighbor distance (µm)	StdDev1stNNDist	–	σ_{r_1}
Error of the standard deviation of the 1st nearest neighbor distance (µm)	ErrStdDev1stNNdist	Error of StdDev1stNNDist, calculated using the jackknife method.	$\Sigma_{\sigma_{r_1}}$
Skewness of the 1st nearest neighbor distance distribution	Skewness1stNNDist	–	$\left\langle \left(\frac{r_1 - \langle r_1 \rangle}{\sigma_{r_1}} \right)^3 \right\rangle$
2nd nearest neighbor mean distance (µm)	2ndNNDist	Average distance from each myelinated axon to the second closest myelinated axon.	$\langle r_2 \rangle$
Standard deviation of the 2nd nearest neighbor distance (µm)	StdDev2ndNNDist	–	σ_{r_2}
Skewness of the 2nd nearest neighbor distance distribution	Skewness2ndNNDist	–	$\left\langle \left(\frac{r_2 - \langle r_2 \rangle}{\sigma_{r_2}} \right)^3 \right\rangle$

TABLE S1. **List of the features considered.** (continued)

3rd nearest neighbor mean distance (μm)	3rdNNDist	Average distance from each myelinated axon to the third closest myelinated axon.	$\langle r_3 \rangle$
Standard deviation of the 3rd nearest neighbor distance (μm)	StdDev3rdNNDist	–	σ_{r_3}
Skewness of the 3rd nearest neighbor distance distribution	Skewness3rdNNDist	–	$\left\langle \left(\frac{r_3 - \langle r_3 \rangle}{\sigma_{r_3}} \right)^3 \right\rangle$
Slope (regression line)	SlopeRegress	Slope of the regression line in the log-log plot of the average nearest neighbor distance $\langle r_n \rangle$ as a function of its rank n, for $n = 8, \ldots, 15$. This slope should be a number close to 0.5 (see Methods for details).	s
Effective local density (μm^{-2})	EffDensity	"Effective" density obtained from the regression line as detailed for the previous quantity (see Methods for details).	ρ_{eff}
Effective local density (with slope=0.5) (μm^{-2})	EffDensitySlope05	Same as previous but now the slope of the regression line is fixed to its expected value 0.5 when performing the fit.	ρ'_{eff}
Mean Hexagonality Index (rad^{-1})	MeanHexagonality	Starting from a central myelinated axon i, we take the sum of the absolute value of the differences of the angles between the line segments joining the centroids of myelinated axons in adjacent Voronoi cells α_{ij} and $\beta = \pi/3$, i.e. $\mathcal{Q}_i = \sum_{j=1}^{\mathcal{N}_i} \lvert \alpha_{ij} - \beta \rvert$. The inverse of $\mathcal{Q}_i + 1$ is the hexagonality index (see Methods for details).	$\langle \Delta \rangle = \left\langle \frac{1}{\mathcal{Q} + 1} \right\rangle$
Standard deviation of Hexagonality Index (rad^{-1})	StdDevHexagonality	–	σ_Δ
Mean number of adjacent Voronoi cells	MeanNNVoronoi	–	$\langle \mathcal{N} \rangle$
Standard deviation of neighbors for Voronoi cells	StdNNVoronoi	-	$\sigma_{\mathcal{N}}$

TABLE S1. **List of the features considered.** (continued)

Pearson R axon area (1st shell)	PearsonR1stShellAxon	Pearson correlation coefficient for the plot of the mean of the areas of the axons in the first neighbor shell (as determined by the Voronoi tessellation) against the area of the central axon.	$R^{(1)}$
Slope axon area (1st shell)	Slope1stShellAxon	Slope of the regression line fit to the plot used in the previous quantity.	$S^{(1)}$
Pearson R axon area (2nd shell)	PearsonR2ndShellAxon	Pearson correlation coefficient for the plot of the mean of the areas of the axons in the second neighbor shell (as determined by the Voronoi tessellation) against the area of the central axon.	$R^{(2)}$
Slope axon area (2nd shell)	Slope2ndShellAxon	Slope of the regression line fit to the plot used in the previous quantity.	$S^{(2)}$
Pearson R Voronoi cell area (1st shell)	PearsonR1stShellVor	Pearson correlation coefficient for the plot of the mean of the areas of the Voronoi cells in the first neighbor shell against the area of the central Voronoi cell.	$\mathcal{R}^{(1)}$
Slope Voronoi cell area (1st shell)	Slope1stShellVor	Slope of the regression line fit to the plot used in the previous quantity.	$\mathcal{S}^{(1)}$
Pearson R Voronoi cell area (2nd shell)	PearsonR2ndShellVor	Pearson correlation coefficient for the plot of the mean of the areas of the Voronoi cells in the second neighbor shell against the area of the central Voronoi cell.	$\mathcal{R}^{(2)}$
Slope Voronoi cell area (2nd shell)	Slope2ndShellVor	Slope of the regression line fit to the plot used in the previous quantity.	$\mathcal{S}^{(2)}$

TABLE S2. **Accuracy and Welch's t-test p-values when considering single features.** The accuracy is calculated using the same methodology as presented in Methods, while its standard deviation is obtained using the five different runs of the cross-validation, as explained in Methods. Since one cannot assume the two age groups to have equal variances, we calculate the Welch's t-test instead of the Student's t-test (n.s. stands for not significant, $p > 0.05$).

Feature name	Accuracy	Welch's p-value	Feature name	Accuracy	Welch's p-value
FracOccupArea	81% ± 2%	$5.6{\times}10^{-10}$	StdDevAbsCurvature	56% ± 2%	0.062 (n.s.)
EffDensity	80% ± 2%	$7.0{\times}10^{-7}$	Skewness	54% ± 1%	0.15 (n.s.)
3rdNNDist	79% ± 5%	$1.0{\times}10^{-8}$	Skewness2ndNNDist	54% ± 3%	0.14 (n.s.)
EffDensitySlope05	76% ± 4%	$2.0{\times}10^{-8}$	StdNNVoronoi	54% ± 1%	0.55 (n.s.)
Density	75% ± 1%	$4.7{\times}10^{-8}$	PearsonR2ndShellAxon	53% ± 2%	0.24 (n.s.)
StdDev3rdNNDist	75% ± 2%	$4.5{\times}10^{-7}$	MeanPerimeter	53% ± 1%	0.82 (n.s.)
NumAxons	74% ± 3%	$5.3{\times}10^{-8}$	Skewness3rdNNDist	53% ± 1%	0.21 (n.s.)
2ndNNDist	74% ± 2%	$2.2{\times}10^{-8}$	StdDevElongation	52% ± 6%	0.016
StdErr1stNNDist	73% ± 2%	$5.5{\times}10^{-6}$	StdDevDiameter	52% ± 3%	0.15 (n.s.)
1stNNDist	70% ± 4%	$1.7{\times}10^{-6}$	Slope1stShellVor	51% ± 2%	0.44 (n.s.)
StdDev2ndNNDist	70% ± 3%	$2.0{\times}10^{-6}$	MeanCircularity	51% ± 6%	0.89 (n.s.)
MeanHexagonality	70% ± 4%	$1.4{\times}10^{-4}$	Skewness1stNNDist	50% ± 2%	0.34 (n.s.)
MeanNNVoronoi	66% ± 3%	0.49 (n.s.)	SlopeRegress	50% ± 2%	0.028
ErrStdDev1stNNdist	64% ± 2%	0.015	MeanBendEnergy	48% ± 1%	0.25 (n.s.)
MeanElongation	61% ± 2%	$6.4{\times}10^{-4}$	StdDevBendEnergy	48% ± 5%	0.94 (n.s.)
StdDev1stNNDist	61% ± 1%	$1.2{\times}10^{-5}$	PearsonR2ndShellVor	48% ± 2%	0.043
Slope1stShellAxon	59% ± 2%	0.12 (n.s.)	StdDevPerimeter	47% ± 3%	0.52 (n.s.)
MeanDiameter	59% ± 4%	0.35 (n.s.)	StdDevHexagonality	46% ± 5%	0.013
Slope2ndShellVor	58% ± 2%	0.067 (n.s.)	MeanAbsCurvature	46% ± 4%	0.011
PearsonR1stShellAxon	57% ± 2%	0.058 (n.s.)	StdDevCircularity	46% ± 1%	0.33 (n.s.)
Slope2ndShellAxon	56% ± 3%	0.27 (n.s.)	PearsonR1stShellVor	42% ± 1%	0.72 (n.s.)
MeanArea	56% ± 2%	0.66 (n.s.)	CoeffVariation	39% ± 1%	0.70 (n.s.)
StdDevArea	56% ± 2%	0.89 (n.s.)			

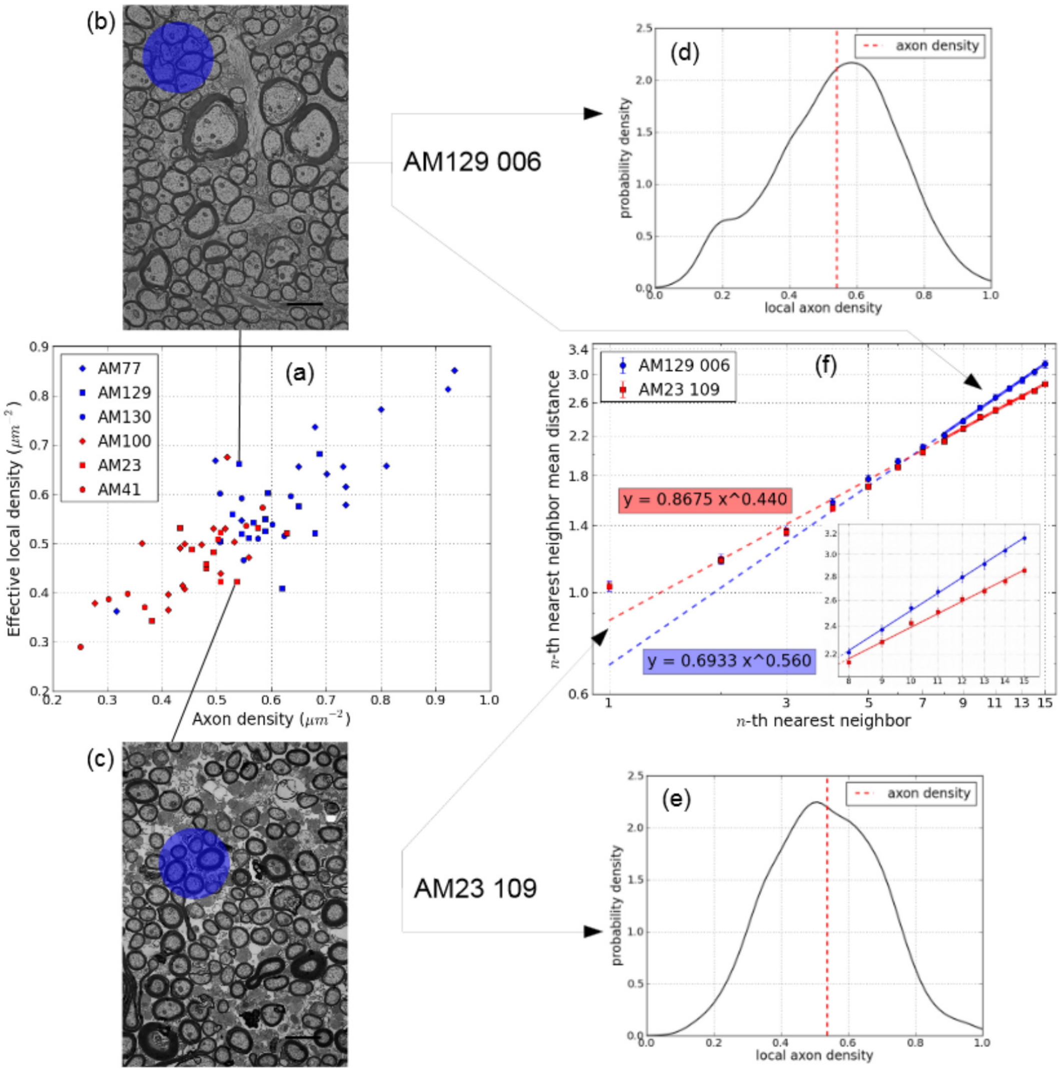


FIG. S1. **Effective Local Density Calculation.** (a) Plot of the effective local density versus the axon density for each sample. (b) and (c) Two different samples with similar axon densities and different effective local densities (AM129 006 and AM23 109) are selected and shown for visual comparison. (d) and (e) Probability density of the local axon densities for the two samples highlighted in figures (b) and (c), calculated via a gaussian kernel density estimation. The local densities were calculated for randomly chosen circular areas of varying radius (an example circle area is highlighted in blue on both samples). The axon density values of each sample are plotted as red dashed lines. (f) Plot of the average n-th nearest neighbor distance in function of the rank n for the two samples highlighted in figures (b) and (c). The full lines are the log-log linear regressions for the points $\mathsf{n} \geq 8$ (see inset figure for a zoomed plot), with corresponding equations in the matching colored boxes. These linear regressions are extrapolated (dashed lines) to $\mathsf{n} = 1$.

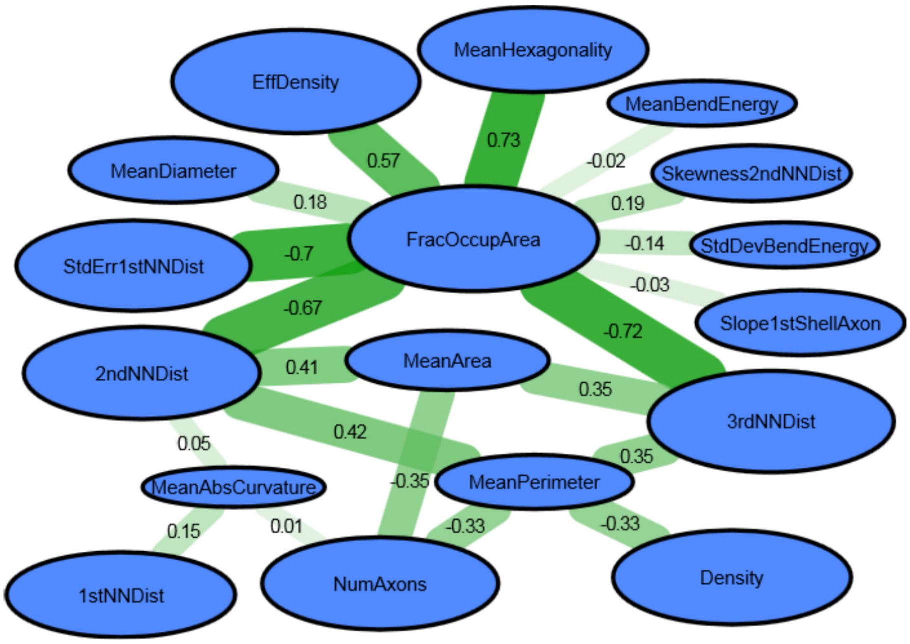
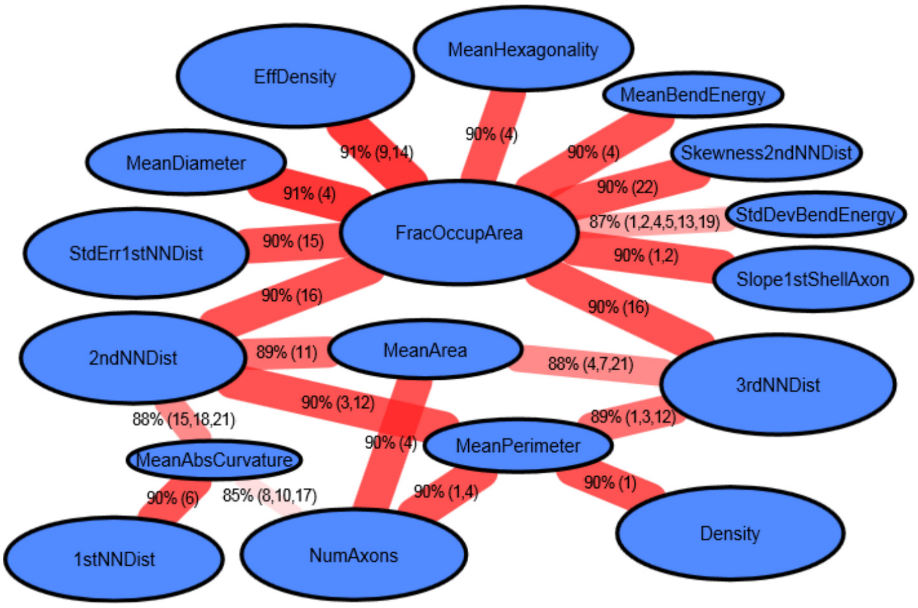


FIG. S2. **Graph of the correlation between features.** The value on each link represents the Pearson correlation coefficient between the features having a link in Fig. 3. Its absolute value is illustrated by both the color and width of the link.



Index	Feature name
1	2ndNNDist
2	3rdNNDist
3	Density
4	EffDensity
5	EffDensitySlope05
6	FracOccupArea
7	MeanAbsCurvature
8	MeanArea
9	MeanDiameter
10	MeanNNVoronoi
11	MeanPerimeter
12	NumAxons
13	Skewness
14	Skewness1stNNDist
15	Skewness2ndNNDist
16	Slope1stShellAxon
17	Slope1stShellVor
18	StdDev1stNNDist
19	StdDev3rdNNDist
20	StdDevBendEnergy
21	StdDevCircularity
22	StdErr1stNNDist

FIG. S3. **Graph of the classification accuracy when using three features.** Each link contains the highest accuracy we can achieve when adding a third feature to each link of the accuracy graph shown in Fig. 3. Shown in parenthesis are the possible indices of this third feature that, when added to the pair, maximize the accuracy. Refer to the table on the right for the name of the feature. We note that all accuracy values shown have an associated error smaller than 2%.

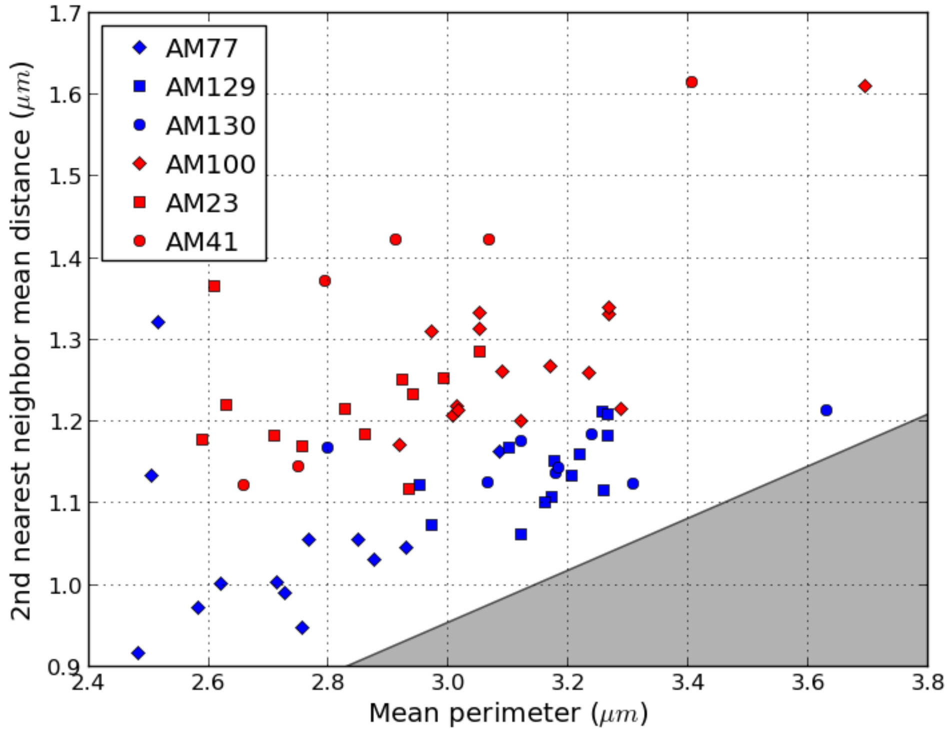


FIG. S4. **Scatter plot of the fornix samples in the second nearest neighbor distance + perimeter space.** These two features provide a good accuracy for age group discrimination and they have a large class scatter distance. Young samples are in blue, while old samples are in red. The grey area represents the expected exclusion region in this two features space when axons are approximated to a set of equal sized disks. In this approximation, the second nearest neighbor distance cannot have a lower value than the diameter of the axon. Young samples have nearest neighbor mean distances much closer to the limit defined by their myelinated axon perimeter values and, therefore, have a more restricted range of values due to the tighter geometric packing of the axons.

SUPPLEMENTARY NOTE 1

Probability distribution of the n-th nearest neighbor mean distance for a homogeneous Poisson point process in d dimensions

Let us consider a homogeneous (i.e. with spatially uniform statistical properties) Poisson point process [55] (i.e. random particle distribution) in $\mathbb{R}^d$, characterized by the uniform mean density $\rho_0 > 0$.

Given a particle of the system at the origin of axes, let us call $\omega_n(r)$ the probability density function (PDF) of the distance of its n-th nearest neighbor, i.e.

$$\omega_n(r)\,dr = \text{Prob}(r \le |\mathbf{r}_n| < r + dr)\ ,$$

where $\mathbf{r}_n$ is the spatial position of the n-th nearest neighbor of the particle located at the origin of the axes. In order to find $\omega_n(r)$ we can generalize the calculation usually adopted for the case $n = 1$. For this system, the positions of different particles are independent and the probability of having exactly N particles in an arbitrary volume V is

$$p(N, V) = \frac{(\rho_0 V)^N}{N!}\, e^{-\rho_0 V}\ , \tag{1}$$

from which it is simple to derive that $\langle N\rangle = \langle N^2\rangle - \langle N\rangle^2 = \rho_0 V$. Using Eq. (1) and considering the mutual independence of the positions of the particles, we can write

$$\text{Prob}(r \le |\mathbf{r}_n| < r + dr) = p(n-1, V(r)) \cdot p(1, S(r)\,dr)\ ,$$

where $V(r)$ is the volume of the d-dimensional sphere of radius r and $S(r)$ is its external surface area so that $S(r)\,dr$ is the volume of the spherical shell between the radii r and $r + dr$. This leads to

$$\omega_n(r)\,dr = \frac{(\rho_0 V(r))^{n-1}}{(n-1)!}\, e^{-\rho_0 V(r)}\, \rho_0 S(r) dr\, e^{-\rho_0 S(r) dr}\ . \tag{2}$$

Note that, since we are interested in the infinitesimal limit for dr, the last exponential in Eq. (2) can be substituted simply by 1. In d dimensions

$$V(r) = \frac{\Omega_d}{d} r^d$$
$$S(r) = \Omega_d r^{d-1}\ ,$$

where Ω_d is the complete solid angle in d dimensions (1 in $d = 1$, 2π in $d = 2$, 4π in $d = 3$, etc.). Therefore, we can finally write

$$\omega_n(r) = \rho_0 \Omega_d r^{d-1} \frac{\left(\rho_0 \frac{\Omega_d}{d} r^d\right)^{n-1}}{(n-1)!} e^{-\rho_0 \frac{\Omega_d}{d} r^d} . \quad (3)$$

As evident from Eq. (2), it is very simple to show that $\omega_n(r)$ is a well normalized PDF, i.e. its integral between 0 and ∞ is equal to 1.

Given Eq. (3), it is simple to calculate any moment $\langle r_n^l \rangle$ of the distance of the n-th nearest neighbor for arbitrary l. In general, in d dimensions and for arbitrary l, the integral will give rise to an appropriate Gamma function.

The n-th nearest neighbor mean distance can be calculated as

$$\langle r_n \rangle = \int_0^\infty r\, \omega_n(r)\, dr = \int_0^\infty dr\, \rho_0 \Omega_d r^d \frac{\left(\rho_0 \frac{\Omega_d}{d} r^d\right)^{n-1}}{(n-1)!} e^{-\rho_0 \frac{\Omega_d}{d} r^d} .$$

By performing the change of variable

$$u = \rho_0 \frac{\Omega_d}{d} r^d ,$$

we can write

$$\langle r_n \rangle = \frac{1}{(n-1)!} \left(\frac{d}{\rho_0 \Omega_d}\right)^{1/d} \int_0^\infty du\, u^{n-1+1/d} e^{-u} .$$

By using the definition of the Euler Gamma function [57] we get the final formula:

$$\langle r_n \rangle = \left(\frac{d}{\rho_0 \Omega_d}\right)^{1/d} \frac{\Gamma(n + 1/d)}{(n-1)!} = \left(\frac{d}{\rho_0 \Omega_d}\right)^{1/d} \frac{\Gamma(n + 1/d)}{\Gamma(n)} . \quad (4)$$

Note that for integer n, $\Gamma(n) = (n-1)!$

In order to get the scaling relation between $\langle r_n \rangle$ and rank n for large n we can use the following expansion for large z of the ratio of two Gamma functions [57]:

$$\frac{\Gamma(z+\alpha)}{\Gamma(z+\beta)} = z^{\alpha-\beta} \left[1 + \frac{(\alpha-\beta)(\alpha+\beta-1)}{2z} + O(z^{-2})\right] . \quad (5)$$

In our case this gives

$$\frac{\Gamma(n + 1/d)}{\Gamma(n)} = n^{1/d} \left[1 - \frac{d-1}{2d^2 n} + O(n^{-2})\right] .$$

Inserting this relation into Eq. (4), we obtain the general formula in d dimensions

$$\langle r_n \rangle = \left(\frac{dn}{\Omega_d \rho_0}\right)^{1/d} \left[1 - \frac{d-1}{2d^2 n} + O(n^{-2})\right] . \quad (6)$$

Specializing our discussion to the case $d = 2$, we have

$$\langle r_n \rangle = \left(\frac{1}{\pi \rho_0} \right)^{1/2} \frac{\Gamma(n + 1/2)}{\Gamma(n)} , \tag{7}$$

where

$$\frac{\Gamma(n + 1/2)}{\Gamma(n)} = n^{1/2} \left(1 - \frac{1}{8n} + O(n^{-2}) \right) .$$

By inserting this relation into Eq. (7), we obtain

$$\langle r_n \rangle = \left(\frac{n}{\pi \rho_0} \right)^{1/2} \left(1 - \frac{1}{8n} + O(n^{-2}) \right) . \tag{8}$$