

## 核孔复合物单分子定位超分辨图像的筛选和重构

侯梦迪<sup>1</sup>, 胡芬<sup>1\*</sup>, 杨建宇<sup>1</sup>, 董浩<sup>1</sup>, 潘雷霆<sup>1,2,3,4\*</sup><sup>1</sup>弱光非线性光子学教育部重点实验室, 南开大学物理科学学院, 泰达应用物理研究院, 天津 300071;<sup>2</sup>药物化学生物学全国重点实验室, 南开大学生命科学学院, 细胞应答交叉科学中心, 天津 300071;<sup>3</sup>南开大学深圳研究院, 广东 深圳 518083;<sup>4</sup>山西大学极端光学协同创新中心, 山西 太原 030006

**摘要** 核孔复合物(NPC)是细胞核膜上由多种蛋白组装而成的复杂结构,在细胞核质交换和信息传递中起着关键作用。单分子定位超分辨成像(SMLM)以其特异性和高成像分辨率成为研究NPC超微结构的主要方法之一。然而,由于抗体标记不完全等因素导致的数据丢失,给后续分析带来了困难。笔者使用SMLM提供的定位信息,结合基于密度的空间聚类算法(DBSCAN)和层次聚类算法进行数据的提取和分类,建立了NPC筛选和定位的分析流程,并采用该处理流程得到了缺失较少且形貌比较均匀的核孔。进一步,基于最小二乘法原理对筛选得到的大量NPC进行质心对齐的重构处理,成功复现出了其经典的八重对称结构,并揭示了核孔蛋白Nup133与Nup98的精确相对位置关系。本研究通过建立核孔筛选和重构的标准流程,填补了SMLM数据的缺失。采用该流程对多种核孔蛋白进行分析,揭示它们的结构特性。所建流程为理解核孔的复杂结构提供了一种高通量的定量分析方法。

**关键词** 生物光学; 超分辨成像; 单分子定位; 核孔复合物; 聚类算法; 重构

中图分类号 O436

文献标志码 A

DOI: 10.3788/CJL231072

## 1 引言

核孔复合物(NPC)是细胞结构中最复杂的生命结构之一,是由30多种不同类型的核孔蛋白(Nup)组成的总分子质量高达110~125 MDa(1 Da=1 u)的超大复合物<sup>[1-3]</sup>,一直被认为是结构生物学的“圣杯”。NPC在细胞的生命活动中扮演着举足轻重的角色,充当着细胞核与胞质之间的通道,调控着RNA、蛋白质等生物分子的运输和交流<sup>[4-5]</sup>。准确解析NPC的结构对于深入理解细胞核功能以及许多重要生命过程的机制至关重要。早在20世纪,科学家们使用电子显微镜<sup>[6-7]</sup>和原子力显微镜<sup>[8]</sup>就已经发现了NPC的八重对称结构。最近,研究人员通过冷冻电镜<sup>[9-13]</sup>揭示了NPC的组成、组装以及基本结构,为后续深入探索NPC的功能和作用打下了坚实的结构基础。

单分子定位超分辨显微(SMLM)技术是一种能够特异性标记目标蛋白并具有超高分辨率的光学成像技术<sup>[14-16]</sup>,有效弥补了电子显微镜在原位成像研究中的不足,已被广泛应用于复杂生物结构的解析。近年来,科学家利用SMLM技术成功解析了黏着斑<sup>[17]</sup>、伪

足小体<sup>[18]</sup>和出芽酵母菌着丝点<sup>[19]</sup>等复杂结构。NPC的直径约为130 nm<sup>[20]</sup>,20 nm成像分辨率的单分子定位超分辨显微镜是研究其结构的理想工具。然而,在长时间成像过程中,可能会出现由稀疏闪烁导致的数据缺失以及动态生命活动造成的成像结果的异质性,给数据分析带来了一定困难。为解决这个问题,图像重建算法应运而生。Szymborska等<sup>[21]</sup>利用粒子平均法分析核孔的SMLM图像,并通过像素图像的分割、对齐和筛选,成功重建了Nup107-160亚复合体结构上Nup的形貌,揭示了它们在NPC中的排列方式,为争议许久的排列提供了新的证据。此后,研究人员进一步发展了改进后的平均算法,实现了目标蛋白的三维重建<sup>[22-23]</sup>。值得一提的是,SMLM数据含有丰富的定位坐标,Curd等<sup>[24]</sup>通过分析SMLM数据的相对位置分布,成功拟合出Nup107的超微结构。这些研究表明,发展新的基于坐标的分析方法可以保留高精度的定位信息,从而可以获取更准确、更详细的信息。

聚类算法是一种强大的工具,用于定量描述和分析SMLM数据,可以对SMLM数据进行提取、分类和解释。常见的聚类算法包括根据密度进行分类的

收稿日期: 2023-08-01; 修回日期: 2023-08-24; 录用日期: 2023-09-04; 网络首发日期: 2023-09-19

基金项目: 国家重点研发计划(2022YFC3400600)、国家自然科学基金(32227802, 12174208)、广东省基础与应用基础研究重大项目(2020B0301030009)、南开大学中央高校基本科研业务费专项资金(2122021337, 2122021405)、高等学校学科创新引智计划(B23045)

通信作者: \*plt@nankai.edu.cn; \*\*hufen@nankai.edu.cn

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)算法<sup>[25]</sup>、根据邻近位点的角度进行分类的 CDC (Clustering algorithm using the local Direction Centrality)算法<sup>[26]</sup>、根据空间分布进行划分的泰森多边形聚类算法<sup>[27]</sup>、以及根据样本点之间的距离进行划分的 K-means 算法<sup>[28]</sup>和层级聚类算法<sup>[29]</sup>等。目前,这些聚类方法主要被用于分析膜蛋白纳米簇<sup>[30-31]</sup>以及两种蛋白的共定位程度<sup>[32-34]</sup>,尚未见到将其应用于亚细胞结构重建的公开报道。NPC 具有独特的簇状分布结构特点,这使得聚类方法非常适合用于对 NPC 的结构进行解析。鉴于此,笔者开发了一套结合多种聚类算法对 NPC 的 SMLM 数据进行分析的流程。该流程旨在筛选出缺失较少且形貌较为均匀的 NPC,并对这些 NPC 进行高通量的统计分析和重建处理,弥补超分辨数据的缺失,揭示 NPC 更加精细的结构信息。

## 2 原理与方法

### 2.1 U2OS 的 NPC 成像

将 U2OS 细胞从培养箱中取出,对其进行多聚甲

醛固定、封闭液(含有 0.1% Triton X-100 和 3% BSA 的 PBS 溶液)渗透、抗体(Nup133 以及 Nup98)孵育后,加入成像缓冲液封片准备进行超分辨成像。制备好的样品安装在自主搭建的 SMLM 成像系统的载物台上<sup>[35]</sup>,选择合适的视野采集 50000 帧图像,对采集的图像进行定位和校正后得到 NPC 的 SMLM 图像。

### 2.2 DBSCAN 算法和层次聚类算法的原理

DBSCAN 算法<sup>[25]</sup>的原理如图 1(a)所示。给定一个样本集,如果一个样本点周围(以该样本点为圆心,以  $E_{ps}$  为半径的圆内)的样本数量超过 MinPts 个,则该样本点被称为核心点,如红色样本点。如果一个样本点周围(以该样本点为圆心,以  $E_{ps}$  为半径的圆内)的样本数量少于 MinPts 个,但是包含核心点,则该样本点被划分为边界点,如黄色样本点。不满足核心点以及边界点条件的样本点被划分为噪声,如蓝色样本点。该算法的核心思想是根据样本点的密度信息进行聚类,其主要优势在于能够自动识别任意形状的簇,并可以分类出离群点,最重要的是不需要预设聚类个数,减少了人为干涉。但是当点分布不均匀时,该算法容易将稀疏分布的簇误识别为噪声。

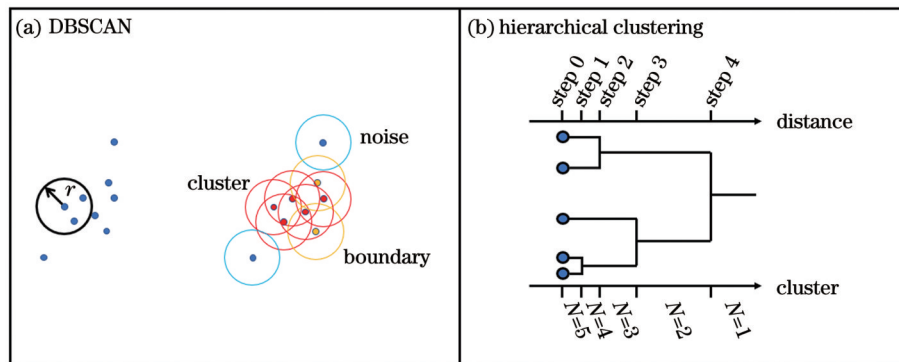


图 1 DBSCAN 和层次聚类的原理示意图。(a)DBSCAN 算法;(b)层次聚类算法

Fig. 1 Principles of DBSCAN and hierarchical clustering. (a) DBSCAN; (b) hierarchical clustering

层次聚类算法<sup>[29]</sup>的原理如图 1(b)所示。该算法先将每个样本点作为一个初始聚类,对于每对样本点,计算它们之间的距离。将距离最近的两个样本点合并为一个新的聚类,形成一个新的节点。之后一直重复该步骤,不断合并距离最近的两个聚类,一直到所有样本点或聚类都被合并成一个根节点。根据所设定的簇值,通过切割聚类树将树划分为聚类簇。该算法能够生成一个聚类树,从而可以提供数据点之间相对关系和聚类之间层次关系的信息,并且能够处理不同大小和形状的聚类,不受限于特定的聚类形状。但是该算法的计算复杂度较高,并且需要预设所需要的簇值。

### 2.3 NPC 数据处理流程

笔者基于目前常用的聚类方法,发展了一种对 NPC 超分辨图像进行筛选与重构的流程,如图 2 所示。首先,利用 DBSCAN 算法(第一次)去除背景,并识别

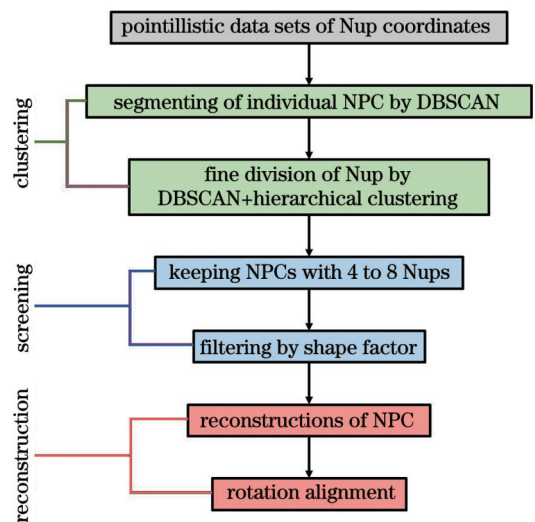


图 2 NPC SMLM 数据处理流程图

Fig. 2 Flowchart of NPC SMLM data processing

出单个 NPC;接着,采用 DBSCAN 结合层次聚类的方法,识别每个 NPC 的 Nup 成分;然后,通过筛选选择含有 4~8 个 Nup 的 NPC,并根据剩余 NPC 的形状因子进行第二轮筛选,保留形状比较均匀的 NPC;接着,将所有 NPC 的质心对齐,实现 Nup 的重构;最后,利用最小二乘法对 Nup 以 NPC 质心为轴进行旋转,重构出具有八重对称结构的 Nup 分布图。

### 3 结果与讨论

#### 3.1 NPC 的筛选

NPC 的 SMLM 图像可能会因为抗体标记不全或者核孔受到挤压等而不完整或者发生变形,这些因素会引入误差,影响对 NPC 数据的定量分析。因此,在 NPC 数据处理的第一步,需要对超分辨数据进行筛选。

笔者模拟了具有一定缺失以及形状扰动的八重对

称 NPC 结构[如图 3(a)所示],并对模拟数据进行了分析。该模拟数据包含 16 个 NPC,其中有 3 个 NPC 含有不足 4 个 Nup,如蓝色方框所示。NPC 以其独特的由小簇组成大簇的结构使得单次聚类划分不能得到全部信息,因此需要进行多次划分定位。首先,为了准确定位整个 NPC,笔者选择使用能够去除噪点的 DBSCAN 算法。DBSCAN 是一种无须预先指定类簇数量的点云数据聚类算法,它根据密度对点云数据进行划分。通过应用 DBSCAN 算法能够有效划分出整个 NPC,并确定其质心位置[如图 3(b)中的蓝色星号所示]。接下来,为了更加准确地划分每个 NPC 中的 Nup,并且不丢失任何信号点,笔者采用了 DBSCAN+层次聚类的方法<sup>[36]</sup>。在第二次应用 DBSCAN 之后,将识别的 Nup 的个数作为输入进行层次聚类,以便在保留所有信号点的情况下进行分簇,并定位 Nup 质心[如图 3(b)中的黑色加号所示]。

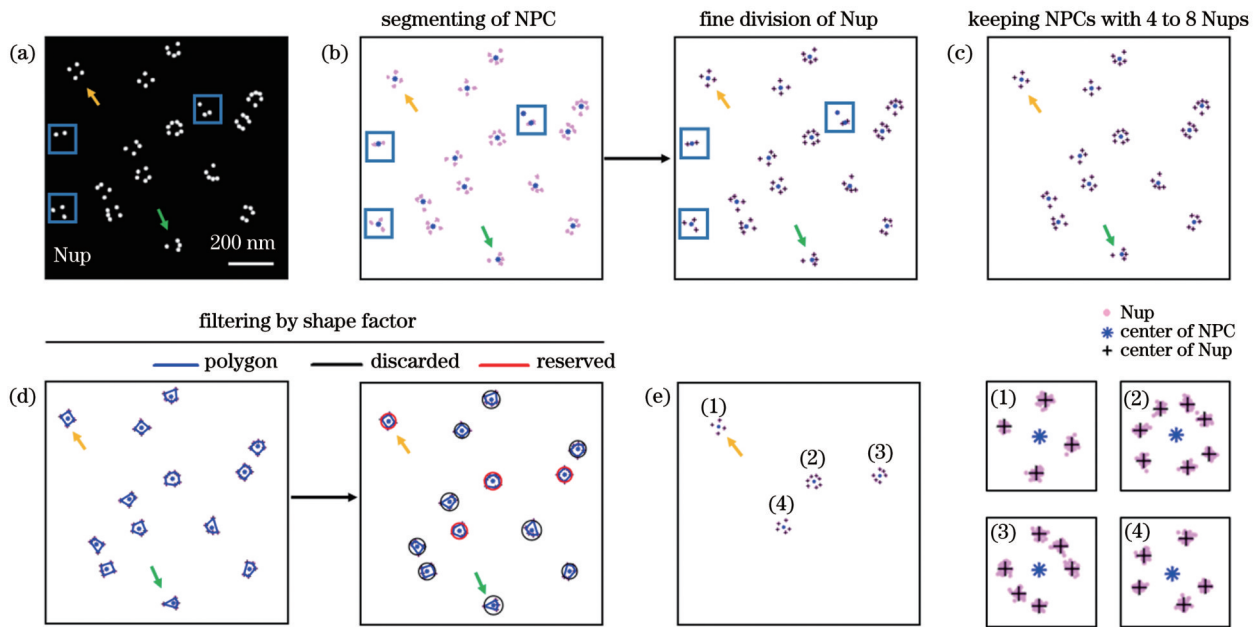


图3 NPC 的筛选。(a)模拟的 NPC 分布图;(b)NPC 的分割以及 Nup 的分割;(c)保留含有 4~8 个 Nup 的 NPC;(d)根据轮廓多边形的形状因子筛选 NPC;(e)满足筛选条件的 NPC

Fig. 3 Screening of NPC. (a) Simulated distribution map of NPCs; (b) segmentation of NPCs and fine division of Nups; (c) keeping NPCs with 4 to 8 Nups; (d) filtering NPCs by shape factor of outline polygons; (e) finally selected NPCs that meet the screening criteria

首先去除少于 4 个 Nup 的 NPC[如图 3(c)所示],然后将定位的所有 Nup 的质心连接形成 NPC 的轮廓多边形[如图 3(d)所示]。定义形状因子  $f$  是以 NPC 质心为圆心的多边形外接圆与多边形的面积比值,即

$$f = \frac{S_{\text{circumcircle}}}{S_{\text{polygon}}} \quad (1)$$

在理想无缺失以及无变形情况下,每个 NPC 应有 8 个 Nup,其轮廓多边形是对角线长为  $2a$  的正八边形,面积为  $4a^2 \sin 45^\circ$ ,其外接圆面积为  $\pi a^2$ 。正八边形的形状因子  $\pi a^2 / (4a^2 \sin 45^\circ)$  约为 1.1。当剩余离散的 4 个 Nup 没有变形时(黄色箭头),其轮廓多边形为

正方形,面积因子  $\pi a^2 / (2a^2)$  约为 1.6。如绿色箭头所示,同样缺失 4 个 Nup,但因其中的 3 个 Nup 接近,其轮廓多边形面积相对于外接圆面积较小,形状因子偏大。因此,可以通过设置面积阈值来进行筛选,保留形状因子小于所设定面积阈值的核孔。形状因子越小,表示缺失的 Nup 越少,Nup 的分布越离散且其形状越接近圆形。当核孔受到挤压而变形时,其形状可能变为椭圆形,导致离散的 4 个 Nup 的轮廓多边形为菱形,此时可以适当调高面积阈值至 2.0(或其他数值,越小表示筛选条件越严格),允许保留部分稍微变形的 NPC。

高于所设定面积阈值的被舍弃的 NPC 用黑色圆圈标记,而低于面积阈值的被保留的 NPC 用红色圆圈标记,如图 3(d)所示。经过筛选,共留下了 4 个 NPC,它们均含有 4 个以上 Nup,且 Nup 的形状都比较圆润,如图 3(e)所示。接下来将对保留下来的 NPC 进行重构分析。

### 3.2 NPC 八重对称结构重构

第一次应用 DBSCAN 所划分的簇代表整个 NPC,其质心代表 NPC 最中间的位置,第二次应用 DBSCAN 并结合层次划分出 Nup 小簇,其质心位置为 Nup 位点(center of Nup)。图 4(a)是含有 7 个 Nup 的 NPC 示意图,将 NPC 的质心置于坐标原点,并建立标

准的 Nup 分布方位(黑色虚线)。连接 Nup 与坐标原点,此时连线与标准方位的较小夹角为  $d\theta$ ,计算每个 Nup 相对于标准方位的角度差平方和,即

$$C_{\theta} = (d\theta_1)^2 + (d\theta_2)^2 + \dots + (d\theta_n)^2, \quad (2)$$

之后将 Nup 以坐标原点为中心顺时针旋转  $\delta_1$  角度,如图 4(b)所示,此时的 Nup 相对于标准方位的角度差平方和为

$$C_{\theta+\delta_1} = (d\theta_1)^2 + (d\theta_2)^2 + \dots + (d\theta_n)^2. \quad (3)$$

以  $0.01\pi$  为梯度旋转  $2\pi$  后得到一系列  $C$  值,当  $C$  值最小时,  $\delta_i$  为最佳旋转角度。以该角度旋转整个 NPC,即可得到与所建立的标准 NPC 方位最接近的结构,如图 4(c)所示。

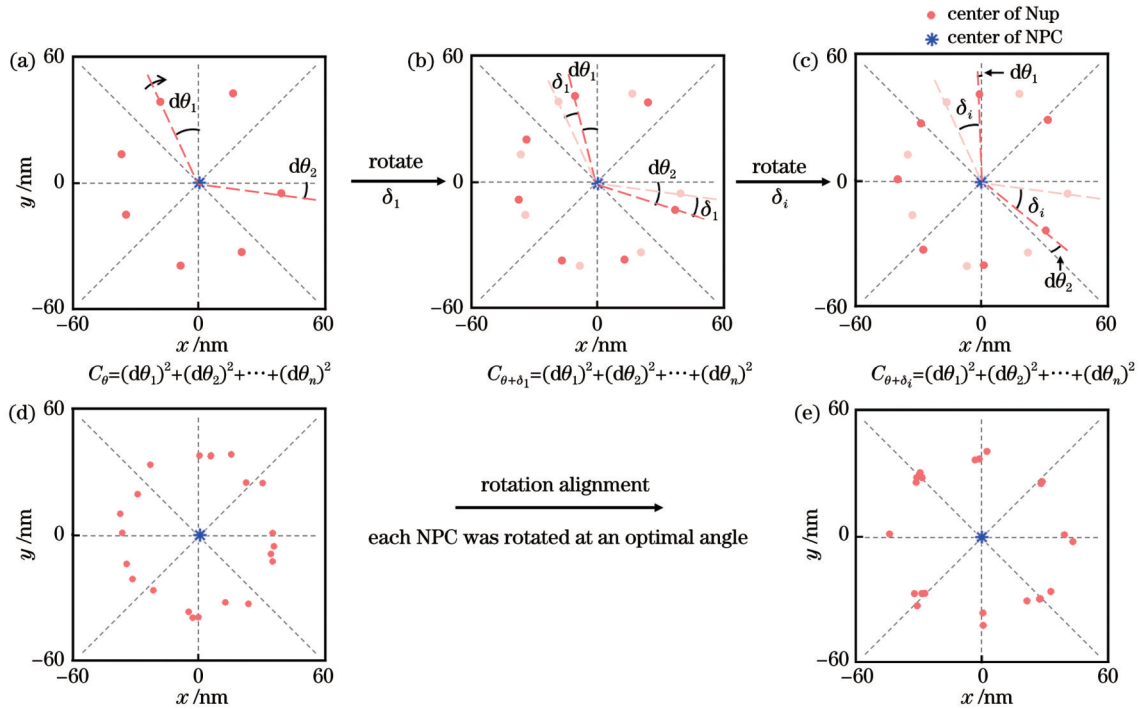


图 4 NPC 的重构。(a)~(c)含有 7 个 Nup 的 NPC 以构建的标准方位旋转的示意图:(a)初始状态;(b)从初始状态开始以 NPC 质心为圆心顺时针旋转  $\delta_1$ ; (c)从初始状态开始以 NPC 质心为圆心顺时针旋转  $\delta_i$ 。(d)筛选出来的所有 NPC 的质心对齐。(e)旋转对齐后的八重对称结构

Fig. 4 Reconstruction of NPC. (a)~(c) A schematic diagram depicting a standard orientation rotation of an NPC composed of 7 Nups: (a) initial state; (b) clockwise rotation  $\delta_1$  around the center of the NPC from initial state; (c) clockwise rotation  $\delta_i$  around the center of the NPC from initial stage. (d) Alignment of all the screened NPC centers. (e) Eightfold symmetric structure after rotational alignment

将 3.1 节中筛选出的所有 NPC 的质心重合并置于坐标原点,即可得到所有 Nup 相对于 NPC 质心的分布[如图 4(d)所示]。重构后的环状结构含有该 Nup 的全部位置信息,其半径为

$$r = \sqrt{(dx)^2 + (dy)^2}, \quad (4)$$

其中,

$$dx = x_{\text{center of nup}} - x_{\text{center of NPC}}, \quad (5)$$

$$dy = y_{\text{center of nup}} - y_{\text{center of NPC}}. \quad (6)$$

通过叠加数百个经筛选后的 NPC,可以获得精度较高的 Nup 的半径统计结果。之后根据已设定的标准结

构,按照每个 NPC 的最佳角度旋转,即可复现出八重对称的 NPC 分布,如图 4(e)所示。

至此,确定了核孔超分辨图像的处理流程,包括划分、筛选以及重构三部分。该流程可以保留形状较为均一的 NPC,选取这些 NPC 进行重构分析可以减小统计误差,得到较为准确的形貌信息。

### 3.3 在实验数据中的应用

选取了两种代表性核孔蛋白 Nup133 和 Nup98 进行分析,其中 Nup133 作为特征性“Y”形结构的组分蛋白被广泛研究<sup>[37]</sup>。

通过 SMLM 对 U2OS 细胞样本的 Nup133 进行标

记[如图 5(a)所示],对 SMLM 图像进行局部放大可以观测到标记了 Nup133 的 NPC 大部分标记不完整,但可以分辨基本轮廓和每个 Nup133 的小簇,如图 5(b)所示。按照前面的流程通过两次定位将含有 4~8 个 Nup133 的 NPC 保留下来[如图 5(c)所示],而后挑选出形状因子在所设定的面积阈值(2.0)以下的核孔,最终在所选区域筛选出若干符合条件的核孔,如图 5(d)所示。

经过多次超分辨成像实验,笔者从 10329 个标记了 Nup133 的 NPC 中共筛选出 558 个符合筛选标准的 NPC 结构,总筛选率约为 5%(见表 1)。筛选出的大部

分 NPC 有 4~7 个 Nup133,很少出现完整标记 8 个 Nup133 的 NPC,如图 5(e)所示。将所有 Nup133 相对于 NPC 质心的位置叠加,叠加后的 Nup133 呈现出标准的环状结构,如图 5(f)所示。通过高斯拟合统计所有 Nup133 的半径,得到统计曲线峰值对应的横坐标为  $(58.4 \pm 0.1)$  nm[如图 5(g)所示],与粒子平均法计算的抗体标记情况下 Nup133 的半径  $(59.4 \pm 0.2)$  nm 非常接近<sup>[21]</sup>。说明所提筛选与重构方法具有较高的精度。高斯拟合的半峰全宽(FWHM)为  $(24.2 \pm 0.4)$  nm,  $\sigma_{133}$  约为 10.3 nm。最后,经过旋转对齐,重构出了 Nup133 的八重对称结构,如图 5(h)所示。

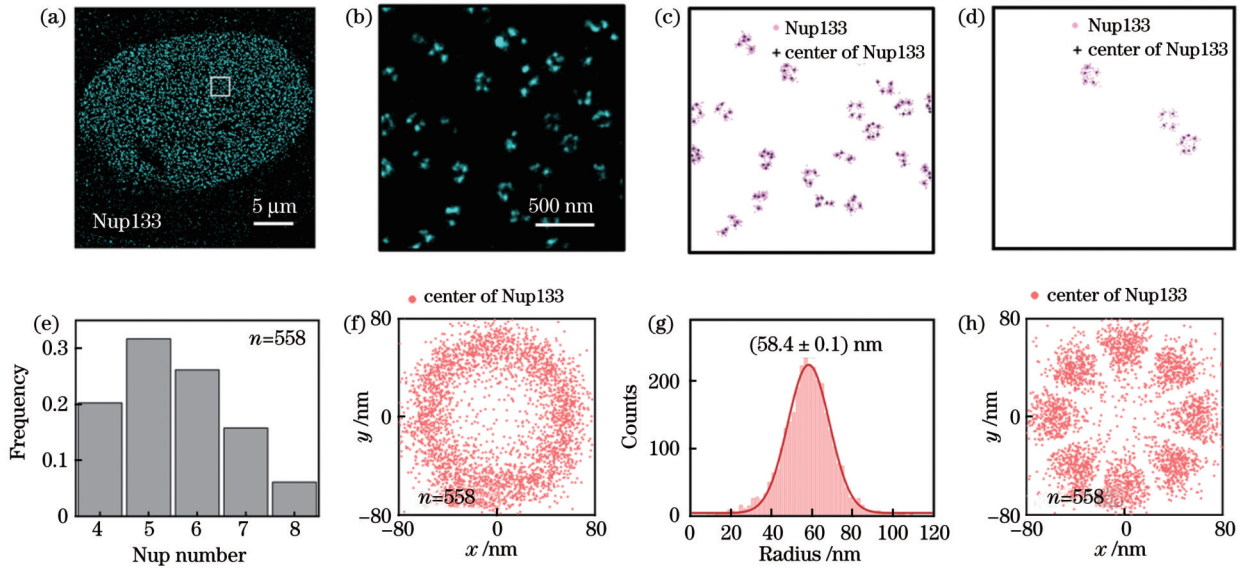


图 5 Nup133 SMLM 图像的重构分析。(a)(b)U2OS 细胞中 Nup133 的代表性 SMLM 图像及其局部放大图;(c)初筛后保留下来的含有 4~8 个 Nup133 的 NPC;(d)最终筛选出的满足条件的 NPC;(e)Nup133 数量分布直方图;(f)对齐所有 NPC 质心的 Nup133 的空间分布;(g)所有 Nup133 的半径分布,其中红色线代表高斯拟合的分布曲线,其峰值对应的横坐标为拟合的 Nup133 半径;(h)旋转对齐后的 Nup133 的空间分布

Fig. 5 Reconstruction and analysis of Nup133 SMLM images. (a)(b) Typical SMLM image of Nup133 in U2OS cells and its zoomed view; (c) remained NPCs with 4-8 Nup133s after preliminary screening; (d) finally selected NPCs in accordance with the screening criteria; (e) histogram of Nup133 number distribution; (f) spatial distribution of Nup133 positions aligned to NPC centers; (g) distribution of Nup133 radius, in which the red line represents the Gaussian-fitted distribution curve and its horizontal ordinate of the peak is the fitted Nup133 radius; (h) spatial distribution of Nup133 after rotational alignment

表 1 Nup133 和 Nup98 的两次筛选比例

Table 1 Ratios of selected Nup133 and Nup98 after twice screening

NuP	A (total NPC number)	B (the number of NPCs with 4 to 8 Nups)	B/A	C (final NPC number)	C/B	C/A
Nup133	10329	3076	30%	558	18%	5%
Nup98	10668	5365	50%	1126	20%	10%

此外,笔者还使用 SMLM 对标记了 Nup98 的 NPC 进行了观测。Nup98 位于 NPC 的内圈<sup>[38]</sup>。结果显示,Nup98 的分布相比 Nup133 更加紧凑,如图 6(a)、(b)所示。同样,经过两次定位后,保留下含有 4~8 个 Nup98 的 NPC[如图 6(c)所示],而后挑选出形状因子在所设定的面积阈值(2.0)以下的核孔,最终在所选区

域筛选出若干符合条件的核孔,如图 6(d)所示。

对于 Nup98,从 10668 个 NPC 中共筛选出 1126 个满足条件的 NPC 结构,总筛选率约为 10%(表 1),比 Nup133 的筛选率高。统计数据显示,Nup133 和 Nup98 筛选的比例差异主要集中于保留含有 4~8 个 Nup 的 NPC 这一步骤上。对比图 5(b)和图 6(b)可以

发现, Nup133 更为离散, 存在较多的不能划分到 NPC 中的小簇, 增加了第一步 DBSCAN 划分簇的基数, 使得在挑选含有 4~8 个 Nup 的 NPC 时去掉了较多小簇, 表现为筛选后的比例降低。

与 Nup133 相似, 大部分 NPC 由 4~7 个 Nup98 定位, 如图 6(e) 所示。将所有 Nup98 相对于 NPC 质心的位置叠加, Nup98 同样呈现出标准的环状结构分布, 如图 6(f) 所示。通过高斯拟合统计所有 Nup98 的半径, 得到统计曲线峰值对应的横坐标为  $(39.7 \pm 0.2)$  nm, 比 Nup133 半径小 18.7 nm, 高斯拟合的 FWHM 为

$(20.3 \pm 0.5)$  nm,  $\sigma_{98}$  约为 8.6 nm。这说明 Nup98 比 Nup133 更靠近 NPC 中间位置。最后, 通过旋转对齐重构出了 Nup98 的八重对称结构, 如图 6(h) 所示。Nup133 和 Nup98 在核孔中的位置如图 6(i) 所示。将 Nup133 重构的 NPC 与 Nup98 重构的 NPC 重合 [如图 6(j) 所示], 可以得到两种蛋白的相对位置关系。核孔复合物是由几十种蛋白组成的复杂结构, 通过解析它们之间的位置关系, 可以更好地了解它们在物质运输过程中的相互作用和功能, 对于理解核孔复合物的生命活动具有重要意义。

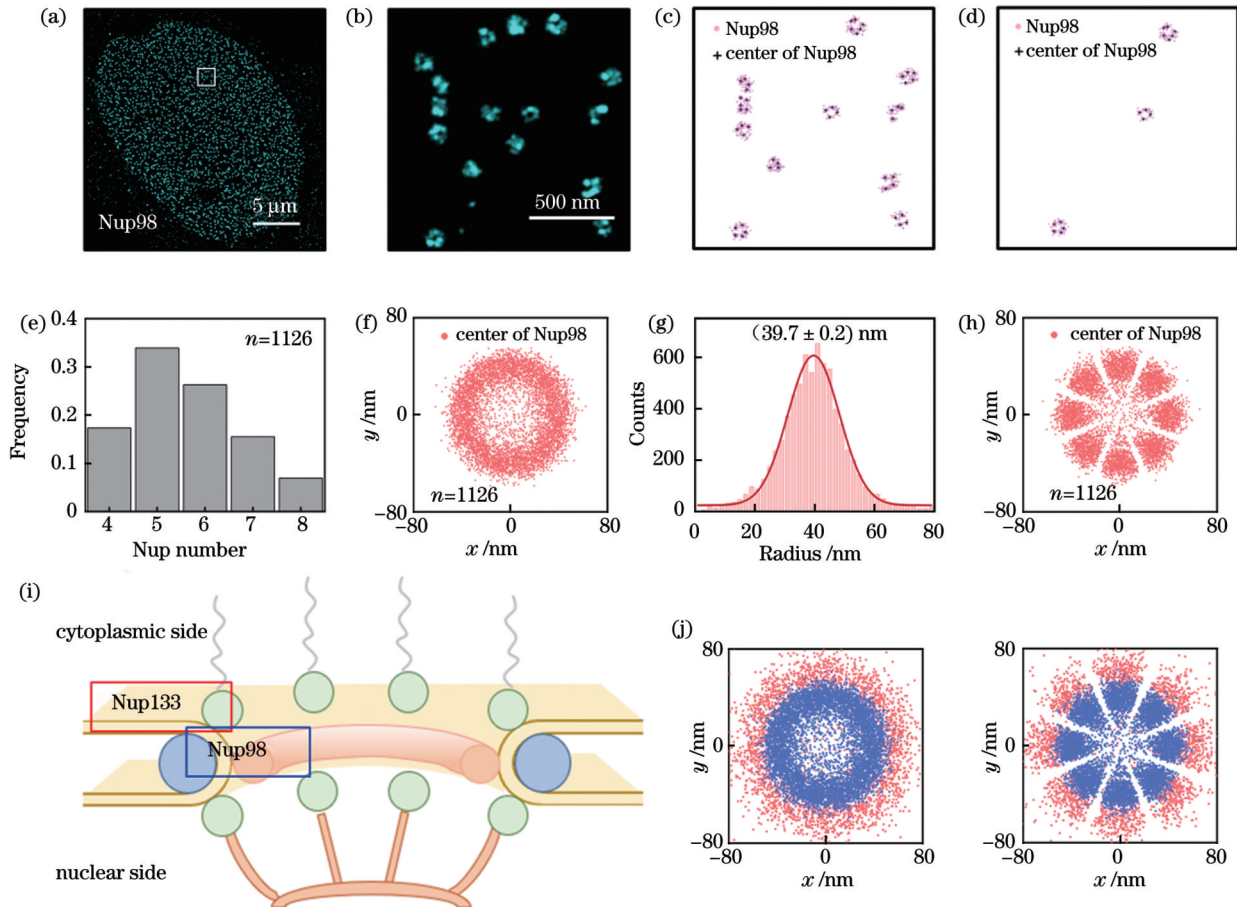


图 6 Nup98 SMLM 图像的重构分析。(a)(b) U2OS 细胞中 Nup98 的代表性 SMLM 图像及其局部放大图;(c) 初筛后保留下来的含有 4~8 个 Nup98 的 NPC;(d) 最终筛选出的满足条件的 NPC;(e) Nup98 数量分布直方图;(f) 对齐所有 NPC 质心的 Nup98 的空间分布;(g) 所有 Nup98 的半径分布, 其中红色线代表高斯拟合的分布曲线, 其峰值对应的横坐标为拟合的 Nup98 半径;

(h) 旋转对齐后的 Nup98 分布;(i) NPC 模型图;(j) Nup133 和 Nup98 分布的叠加

Fig. 6 Reconstruction and analysis of Nup98 SMLM images. (a)(b) Typical SMLM image of Nup98 in U2OS cells and its zoomed view; (c) remained NPCs with 4-8 Nup98s after preliminary screening; (d) finally selected NPCs in accordance with the screening criteria; (e) histogram of Nup98 number distribution; (f) spatial distribution of Nup98 positions aligned to NPC centers; (g) distribution of Nup98 radius, in which the red line represents the Gaussian-fitted distribution curve and its horizontal ordinate of the peak is the fitted Nup98 radius; (h) the distribution of Nup98 after rotational alignment; (i) model diagram of NPC; (j) overlay of Nup133 and Nup98 distributions

## 4 结 论

针对核孔复合物的单分子定位超分辨图像, 笔者提出了一种基于聚类方法的筛选和重构的处理流程。该流程主要包括分类、筛选和重构三部分。通过两次

聚类识别 NPC 以及 Nup 成分, 筛选含有 4~8 个 Nup 的形貌均匀的 NPC, 并进行重构分析, 最终通过旋转成功重构出具有八重对称结构的 NPC。将所提方法应用于 Nup133 和 Nup98, 实验结果显示: Nup133 的半径为  $(58.4 \pm 0.1)$  nm, 与粒子平均法计算的半径接近;

Nup98 的半径为  $(39.7 \pm 0.2)$  nm, 比 Nup133 的半径小 18.7 nm, 表明 Nup98 更靠近核孔的中间位置。这些结果表明所提方法能够重构 NPC 并复现其八重对称结构, 给出了准确的定位信息, 有助于深入了解核孔复合物的组成。这种利用聚类分析重建亚细胞结构的方法也可以推广至其他类核孔结构, 如中心粒、鞭毛基体等, 或者其他具有各向同性对称的结构, 为解析复杂生命结构提供了重要策略和方法。

### 参 考 文 献

- [1] Tai L H, Yin G L, Sun F, et al. Cryo-electron microscopy reveals the structure of the nuclear pore complex[J]. *Journal of Molecular Biology*, 2023, 435(9): 168051.
- [2] Allegretti M, Zimmerli C E, Rantos V, et al. In-cell architecture of the nuclear pore and snapshots of its turnover[J]. *Nature*, 2020, 586(7831): 796-800.
- [3] Lin D H, Hoelz A. The structure of the nuclear pore complex (an update)[J]. *Annual Review of Biochemistry*, 2019, 88: 725-783.
- [4] Schreiner S M, Koo P K, Zhao Y, et al. The tethering of chromatin to the nuclear envelope supports nuclear mechanics[J]. *Nature Communications*, 2015, 6: 7159.
- [5] Sakuma S, D'Angelo M A. The roles of the nuclear pore complex in cellular dysfunction, aging and disease[J]. *Seminars in Cell & Developmental Biology*, 2017, 68: 72-84.
- [6] Alber F, Dokudovskaya S, Veenhoff L M, et al. The molecular architecture of the nuclear pore complex[J]. *Nature*, 2007, 450(7170): 695-701.
- [7] Rout M P, Aitchison J D, Suprpto A, et al. The yeast nuclear pore complex: composition, architecture, and transport mechanism [J]. *The Journal of Cell Biology*, 2000, 148(4): 635-651.
- [8] Stoffer D, Goldie K N, Feja B, et al. Calcium-mediated structural changes of native nuclear pore complexes monitored by time-lapse atomic force microscopy[J]. *Journal of Molecular Biology*, 1999, 287(4): 741-752.
- [9] Bley C J, Nie S, Mobbs G W, et al. Architecture of the cytoplasmic face of the nuclear pore[J]. *Science*, 2022, 376(6598): eabm9129.
- [10] Petrovic S, Samanta D, Perriches T, et al. Architecture of the linker-scaffold in the nuclear pore[J]. *Science*, 2022, 376(6598): eabm9798.
- [11] Mosalaganti S, Obarska-Kosinska A, Siggel M, et al. AI-based structure prediction empowers integrative structural analysis of human nuclear pores[J]. *Science*, 2022, 376(6598): eabm9506.
- [12] Fontana P, Dong Y, Pi X, et al. Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold[J]. *Science*, 2022, 376(6598): eabm9326.
- [13] Zhu X C, Huang G, Zeng C, et al. Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex[J]. *Science*, 2022, 376(6598): eabl8280.
- [14] Rust M J, Bates M, Zhuang X W. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)[J]. *Nature Methods*, 2006, 3(10): 793-796.
- [15] Betzig E, Patterson G H, Sougrat R, et al. Imaging intracellular fluorescent proteins at nanometer resolution[J]. *Science*, 2006, 313(5793): 1642-1645.
- [16] Jungmann R, Avendaño M S, Woehrstein J B, et al. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT[J]. *Nature Methods*, 2014, 11(3): 313-318.
- [17] Kanchanawong P, Shtengel G, Pasapera A M, et al. Nanoscale architecture of integrin-based cell adhesions[J]. *Nature*, 2010, 468(7323): 580-584.
- [18] Hu F, Zhu D L, Dong H, et al. Super-resolution microscopy reveals nanoscale architecture and regulation of podosome clusters in primary macrophages[J]. *iScience*, 2022, 25(12): 105514.
- [19] Cieslinski K, Wu Y L, Nechyporenko L, et al. Nanoscale structural organization and stoichiometry of the budding yeast kinetochore[J]. *Journal of Cell Biology*, 2023, 222(4): 202209094.
- [20] Thevathasan J V, Kahnwald M, Ciesliński K, et al. Nuclear pores as versatile reference standards for quantitative superresolution microscopy[J]. *Nature Methods*, 2019, 16(10): 1045-1053.
- [21] Szymborska A, de Marco A, Daigle N, et al. Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging[J]. *Science*, 2013, 341(6146): 655-658.
- [22] Salas D, le Gall A, Fiche J B, et al. Angular reconstitution-based 3D reconstructions of nanomolecular structures from superresolution light-microscopy images[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(35): 9273-9278.
- [23] Sieben C, Banterle N, Douglass K M, et al. Multicolor single-particle reconstruction of protein complexes[J]. *Nature Methods*, 2018, 15(10): 777-780.
- [24] Curd A P, Leng J, Hughes R E, et al. Nanoscale pattern extraction from relative positions of sparse 3D localizations[J]. *Nano Letters*, 2021, 21(3): 1213-1220.
- [25] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]// *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, August 2-4, 1996, Portland, Oregon. New York: ACM Press, 1996: 226-231.
- [26] Peng D H, Gui Z P, Wang D H, et al. Clustering by measuring local direction centrality for data with heterogeneous density and weak connectivity[J]. *Nature Communications*, 2022, 13: 5455.
- [27] Levet F, Hosity E, Kechkar A, et al. SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data[J]. *Nature Methods*, 2015, 12(11): 1065-1071.
- [28] Pratim M P. Probabilistic optically-selective single-molecule imaging based localization encoded (POSSIBLE) microscopy for ultra-superresolution imaging[J]. *PLoS One*, 2020, 15(11): e0242452.
- [29] Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview[J]. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2012, 2(1): 86-97.
- [30] Pritchard H A T, Pires P W, Yamasaki E, et al. Nanoscale remodeling of ryanodine receptor cluster size underlies cerebral microvascular dysfunction in Duchenne muscular dystrophy[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(41): E9745-E9752.
- [31] Yan Q Y, Lu Y T, Zhou L L, et al. Mechanistic insights into GLUT1 activation and clustering revealed by super-resolution imaging[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(27): 7033-7038.
- [32] Pagoon S V, Nicovich P R, Mollazade M, et al. Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data[J]. *Molecular Biology of the Cell*, 2016, 27(22): 3627-3636.
- [33] Levet F, Julien G, Galland R, et al. A tessellation-based colocalization analysis approach for single-molecule localization microscopy[J]. *Nature Communications*, 2019, 10: 2379.
- [34] Ejdrup A L, Lycas M D, Lorenzen N, et al. A density-based enrichment measure for assessing colocalization in single-molecule localization microscopy data[J]. *Nature Communications*, 2022, 13: 4388.
- [35] Pan L T, Yan R, Li W, et al. Super-resolution microscopy reveals the native ultrastructure of the erythrocyte cytoskeleton[J]. *Cell Reports*, 2018, 22(5): 1151-1158.
- [36] 杨建宇, 胡芬, 邢福临, 等. 结合多次DBSCAN和层次聚类算法的膜蛋白单分子定位超分辨图像分割[J]. *中国激光*, 2023, 50(3): 0307106.
- [37] Yang J Y, Hu F, Xing F L, et al. Clustering segmentation for

single-molecule localization super-resolution image of membrane protein by combining multi-step DBSCAN and hierarchical clustering algorithm[J]. Chinese Journal of Lasers, 2023, 50(3): 0307106.

[37] Walther T C, Alves A, Pickersgill H, et al. The conserved

Nup107-160 complex is critical for nuclear pore complex assembly [J]. Cell, 2003, 113(2): 195-206.

[38] Hoogenboom B W, Hough L E, Lemke E A, et al. Physics of the nuclear pore complex: theory, modeling and experiment[J]. Physics Reports, 2021, 921: 1-53.

## Screening and Reconstruction for Single-Molecular Localization Superresolution Images of Nuclear Pore Complexes

Hou Mengdi<sup>1</sup>, Hu Fen<sup>1\*\*</sup>, Yang Jianyu<sup>1</sup>, Dong Hao<sup>1</sup>, Pan Leiting<sup>1,2,3,4\*</sup>

<sup>1</sup>Key Laboratory of Weak-Light Nonlinear Photonics, Ministry of Education, School of Physics and TEDA Institute of Applied Physics, Nankai University, Tianjin 300071, China;

<sup>2</sup>Frontiers Science Center for Cell Responses, State Key Laboratory of Medicinal Chemical Biology, College of Life Sciences, Nankai University, Tianjin 300071, China;

<sup>3</sup>Shenzhen Research Institute of Nankai University, Shenzhen 518083, Guangdong, China;

<sup>4</sup>Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, Shanxi, China

### Abstract

**Objective** The nuclear pore complex (NPC) is an intricate structure comprising multiple distinct nuclear pore proteins known as nucleoporins (Nups). It plays a crucial role in the transformation of matter and information between the nucleus and cytoplasm. With a total molecular weight of 110–125 MDa, the NPC is hailed as the “holy grail” of structural biology. Scientists have used such techniques as electron microscopy, atomic force microscopy, and cryoelectron microscopy to collectively reveal the composition, assembly, and ultrastructure of the NPC, providing a solid structural foundation for further exploration of its functions. The diameter of the NPC is approximately 130 nm. Therefore, single-molecule localization microscopy (SMLM) with an imaging resolution of 20 nm is an ideal tool for studying the ultrastructure of NPC. However, during long-term imaging, data loss may occur because of sparse blinking, and the dynamic activities of life also lead to heterogeneity in imaging results, posing challenges for data analysis. To address these issues, corresponding image reconstruction methods must be developed. Clustering algorithms are powerful tools for quantitative extraction, classification, and analysis of SMLM data. The unique clustered distribution structure of the NPC makes clustering methods highly suitable for structural analysis of the NPC. Therefore, to compensate for the limitations of SMLM data and obtain more detailed structural information about the NPC, a processing procedure for SMLM images of the NPC was developed in this study based on clustering algorithms. It involves screening out NPC structures with a more uniform morphology, followed by subjecting these structures to high-throughput statistical analysis and reconstruction.

**Methods** After PFA fixation, permeabilization with a blocking buffer, and labeling with antibodies (Nup133 and Nup98), U2OS cells were imaged by a self-built SMLM imaging system. A total of 50000 frames were captured after appropriate fields of view were selected. Through localization and drift correction processes, corresponding SMLM images were obtained. After the regions of interest were selected, the coordinate data with high localization accuracy were preserved for further analysis. First, a first round of density-based spatial clustering of applications with noise clustering (DBSCAN) analysis was used to remove background noise, identify individual NPCs, and determine the centroids of the NPCs (Fig. 3). To achieve a more accurate delineation of each Nup within every NPC in the case of retaining all signal points, a combination of the DBSCAN algorithm and hierarchical clustering was employed in the second round of delineation. In the second round of DBSCAN analysis, the algorithm was applied to identify the number of individual Nups within each NPC, and the data were further input into a hierarchical clustering algorithm for refinement of Nup localization. Subsequently, NPCs containing four to eight Nups were retained, and a second screening based on shape factors was performed to preserve NPCs with more uniform morphologies. Finally, the centroids of all remaining NPCs were aligned to obtain the complete distribution of labeled Nups in the NPCs. Using the least-squares method with NPC centroids as the center, a reconstruction of the Nup distribution with octagonal symmetry was achieved (Fig. 4). The reconstructed structure can be used to analyze the spatial characteristics of the Nup.

**Results and Discussions** Nup133, as a characteristic “Y”-scaffold-shaped component protein, has received extensive attention in recent research. Through statistical analysis of multiple datasets, the first round of the DBSCAN algorithm identified 10329 NPCs (Fig. 5). Among them, 3076 NPCs containing four to eight Nup133 were present, accounting for approximately 30% of the total. By



selecting based on shape factors, a final set of 558 NPCs with relatively regular shape was obtained, accounting for approximately 5% of the total (Table 1). The retained NPCs were aligned by their centroids, resulting in an overlapped NPC image. Gaussian fitting was applied to calculate the radii of all Nup133, with the peak corresponding to a horizontal coordinate of  $(58.4 \pm 0.1)$  nm. This value is very close to the Nup133 radius of  $(59.4 \pm 0.2)$  nm calculated using the particle averaging method with antibody labeling. This further demonstrates the high-precision performance of the screening and reconstruction methods used in this study. In addition, the same analysis process was applied to analyze NPCs labeled with Nup98. Compared with that of Nup133, the distribution of Nup98 located in the inner ring of the NPC is more condensed (Fig. 6). A total of 10668 NPCs were analyzed, and 1126 NPCs were ultimately retained, accounting for approximately 10% of the total (Table 1). Similarly, the remaining NPCs labeled with Nup98 were aligned by the centroids, and Gaussian fitting was applied to the overlapped Nup98, resulting in a peak corresponding to a horizontal coordinate of  $(39.7 \pm 0.2)$  nm (Fig. 6). Compared with that of Nup133, the radius of Nup98 is smaller by 18.7 nm, indicating that Nup98 is closer to the center position of the NPC than Nup133. Finally, the eightfold symmetric structure of Nup133 and Nup98 was successfully reconstructed using the rotation alignment method, which is consistent with the acknowledged model.

**Conclusions** The present study proposes a processing workflow based on clustering methods for screening and reconstruction of SMLM images of the NPC. The workflow has three main parts: classification, screening, and reconstruction. By performing two rounds of clustering to identify the NPC and Nup components, NPCs with a uniform shape containing four to eight Nups are selected and subjected to reconstruction analysis. The NPC with an eightfold symmetric structure is successfully reconstructed using the proposed workflow. Experimental results on Nup133 and Nup98 show that the radius of Nup133 is  $(58.4 \pm 0.1)$  nm, which closely aligns with the radius determined by the particle averaging method. The radius of Nup98 is  $(39.7 \pm 0.2)$  nm, indicating that Nup98 is situated in closer proximity to the central region of the nuclear pore. The proposed method reproduces the eightfold symmetric structure of the NPC, providing accurate localization information and aiding in a deeper understanding of the composition of this important structure. This clustering-based reconstruction method can also be extended to other nuclear pore-like structures, such as centrioles and basal bodies, or other structures with isotropic symmetric features, offering important strategies and methods for deciphering complex biological structures.

**Key words** bio-optics; super-resolution microscopy; single-molecule localization; nuclear pore complexes; clustering algorithms; reconstruction