High-throughput data collection technologies are becoming increasingly common in many fields, especially in biomedical applications involving single cell data (e.g., scRNA-seq and CyTOF). These introduce a rising need for exploratory analysis to reveal and understand hidden structure in the collected (high-dimensional) Big Data. A crucial aspect in such analysis is the separation of intrinsic data geometry from data distribution, as (a) the latter is typically biased by collection artifacts and data availability, and (b) rare subpopulations and sparse transitions between meta-stable states are often of great interest in biomedical data analysis. In this talk, I will show several tools that leverage manifold learning, graph signal processing, and harmonic analysis for biomedical (in particular, genomic/proteomic) data exploration, with emphasis on visualization, data generation/augmentation, and nonlinear feature extraction. A common thread in the presented tools is the construction of a data-driven diffusion geometry that both captures intrinsic structure in data and provides a generalization of Fourier harmonics on it. These, in turn, are used to process data features along the data geometry for denoising and generative purposes. Finally, I will relate this approach to the recently-proposed geometric scattering transform that generalizes Mallat’s scattering to non-Euclidean domains, and provides a mathematical framework for theoretical understanding of the emerging field of geometric deep learning.