1 Introduction

The central dogma of molecular biology, despite its many simplifications, has guided research for more than fifty years. A substantial technological toolbox has been developed to measure the sequence information encoded by theory's biopolymers at all three levels: DNA, RNA, and protein. Moreover, impressive progress has been made in recent years utilizing optical microscopy and the polymerase chain reaction for measurement of amplified colonies of individual nucleic acids. This so-called second generation of sequencing technologies has extend DNA sequencing into the single-molecule realm, permitting high throughput whole-genome measurements of individual cells and tissues, promising concomitant scientific and biomedical payoff. In contrast, the measurement of individual protein molecules has lagged behind, and investigators must rely on ensemble measurements of protein sequence information from many cells, masking cell-to-cell variations, or else measure only the most abundant proteins in single cell measurements, masking the impact of low-copy number proteins. **something something affinity reagents?** A viable single-molecule protein sequencing technology is therefore an attractive research direction.

1.1 State of the art: mass-spectrometry-based protein sequencing

Modern mass spectrometry (MS) equipment enables routine proteome quantification through direct measurement of expressed proteins. This approach exhibits attomole detection sensitivities for whole proteins and subattomole sensitivities after fractionation and stochastic sampling, implying a detection dynamic range spanning four orders of magnitude[13]. However, expression levels for a typical mammalian proteome span seven orders of magnitude, and low (1-1000) copy-number proteins making up approximately 10% of expressed protein species tend to remain undetected by this method. These proteins are important despite their low frequency: at least one in four display a genetic interaction in double knockout experiments. mRNA transcriptomic analysis correlates well with expressed protein levels at high copy-numbers, but is not as useful a proxy in the low copy number / single-molecule regime where stochastic effects dominate.