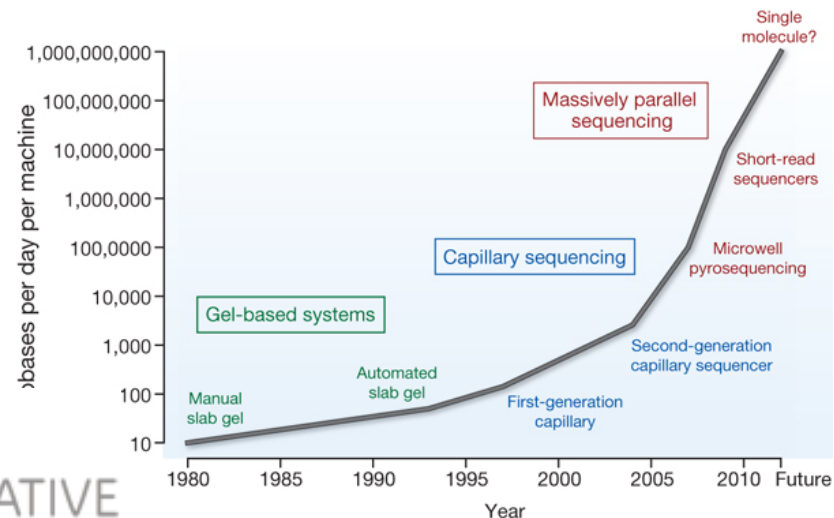


# Taking HEP data management outside HEP

- The Simons Foundation Autism Research Initiative is sequencing the genomes of large numbers of families with autism in order to research the genetic risk factors
- The data rates produced by gene sequencing have increased rapidly over time with the result that the datasets have reached the petabyte scale
- The traditional methods of data management do not work well at this scale



# Taking HEP data management outside HEP

- Many of the data management challenges faced by genome researchers are similar to those already faced by high energy physics experiments
- Fermilab has developed a data management system – SAM – which is used by multiple HEP physics experiments
- A system which is generic enough to handle multiple experiments is also generic enough to handle genomics data
- Cataloging files with metadata will allow SFARI to organize, distribute, and process their data more effectively than at present