**The Wellderly Project Storage & Analytics Requirements**

The Wellderly project uses a collection of advanced aged subjects (85 or older) who have had their genome expressed and analyzed for mutations. These data will form the raw input to the platform that will then use analytic methods to evaluate genetic patterns found.

There are a couple of known goals of this project. The first is to develop a superior reference genome to the current published dataset for human beings known as HG19. Since these subjects live to such an advanced age there is a presumption that they would contain less life threatening mutations than the average or randomly selected humans, which form the basis of the current reference genome. Second, since these subjects have reached an advanced age without major illnesses, we would like to analyze these data to see what genetic characteristics lead to longer life. This is a rich a homogenous dataset from an age perspective anyway and could lead to many other important discoveries that are yet to be anticipated beyond the goals stated above. For example, we will have some demographic data along with these datasets that might provide valuable insight.

The project will initially store the genetic variations of 550 subjects with about 1000 subjects to be added later. The volume of this data for these inputs alone is estimated to be 1 to 2 TB. On top of that, if we were successful at building a newer better version of a reference genome this would take an addition 1 TB of storage. Obviously there will be much more data stored to the output of analytics plus data retrieved from other systems required for the analysis or to fully document our results.