We are proposing a new method to discover unknown genes in any genome. Using our GenomePro framework we process raw genomic input files, of any size, from formats such as NGS, fasta, and GBK, extracting DNA, RNA or Protein sub-sequences of any length. Our work can be applied to any life form.

Human DNA contains approximately 20,000–25,000 known genes, with different genes in different cells getting activated and suppressed as a means of generating a diversity of cells for specific tissues including skin, liver, heart, and others. Of the 3.2 billion bases in the Human Genome we find genes in about 2% of the genome, the remaining 98% is known as dark matter. By examining the dark matter areas we embark in predicting the location of currently unknown genes.

The Encyclopedia of DNA Elements theorizes that at least 80% of the Human DNA serves some biochemical purpose. The new 1000 Genomes Project produces very large genomic files of Petabytes in size, needed to be understood. New technologies and improvements in current technologies allow us to arrive to clinical diagnostics in a shorter time.

Scientists have discovered that essentially all coding and non-coding RNA originates at the same types of locations along the human genome. They have been able to understand how genomic "dark matter"—called non-coding RNA— originates. Since the genetic origins of many diseases reside outside of the coding region of the genome, these findings may also help us to find where complex-disease traits reside.