

# Haplotype-Statistical Models: Integrating to the Conceptual Model of the Human Genome



UNIVERSIDAD  
POLITECNICA  
DE VALENCIA

José Fabián Reyes Román [jreyes@pros.upv.es]  
Doctoral Program in Computer Science  
Software Engineering, Formal Methods and Information Systems



Dr. Óscar Pastor López  
Research Centre on Software Production Methods (PROS)  
Universitat Politècnica de València, Spain

## Introduction

### Motivation:

Resolve 2 problems presented by the current genomic chaos.

- Dispersion / Data Inconsistency: A data set that do not match each other, the great diversity of analyzes performed without a point in common for the proper management.
- High precision / quality in determining data: use approaches and statistical models that allow us to refine the data and thus to generate a reliable and accurate results.

### Contribution:

Improve the management of data in genomic world. The approach is to improve data management by integrating all knowledge. Conceptual models help us to formalize knowledge unambiguously and orderly manner.

### Goal:

Extend CMHG (Conceptual Model of the Human Genome) to thereby represent new knowledge and The progress made in the bioinformatics area, and generate application / more efficient software applications and powerful frameworks.

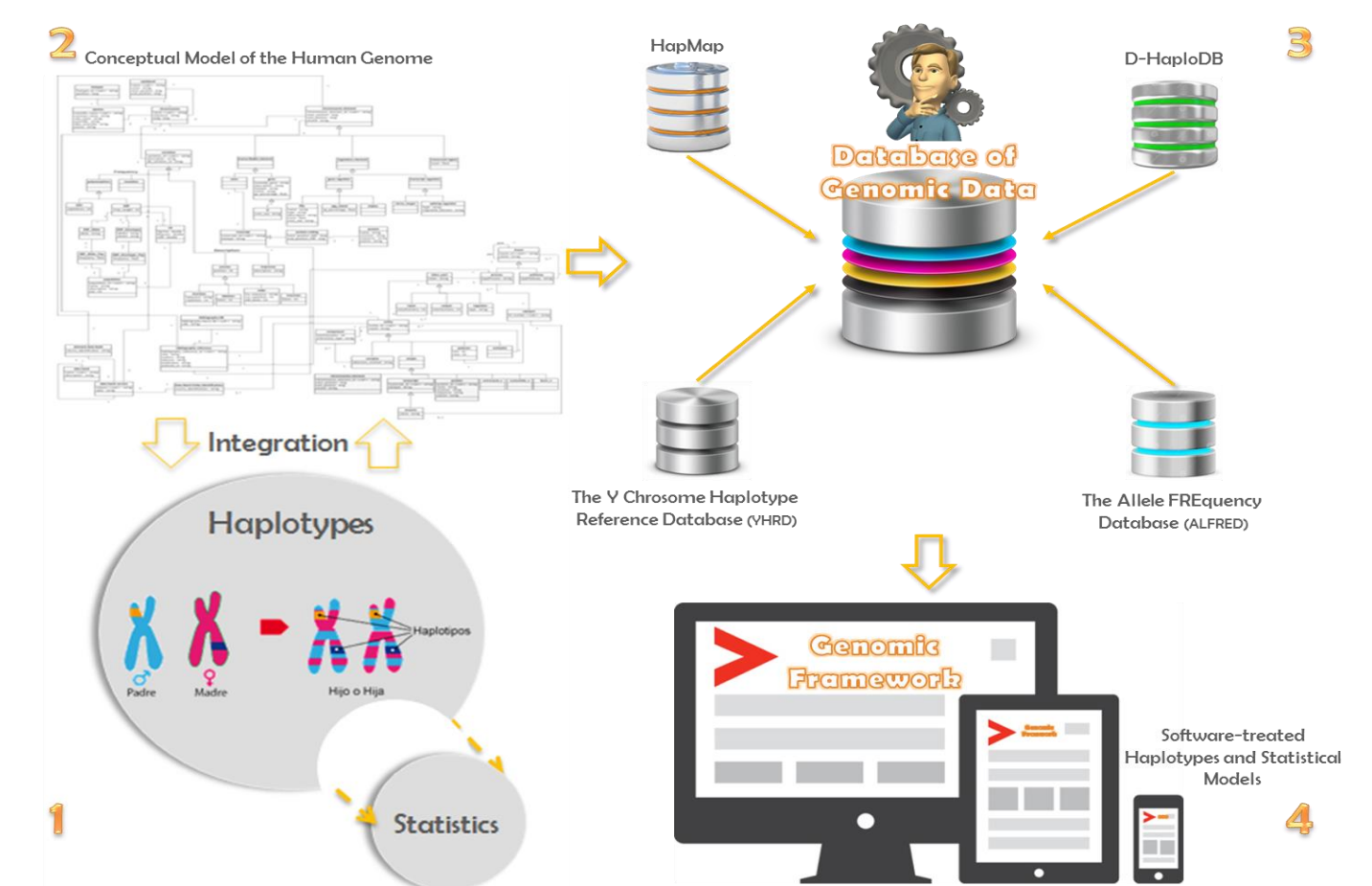
## Objectives

### General:

The overall objective is to address the dispersion and inconsistency of genomic data. As a solution we propose to extend the Conceptual Model of the Human Genome with a criteria untreated up to date: determine the data quality of genomic information.

### Specifics:

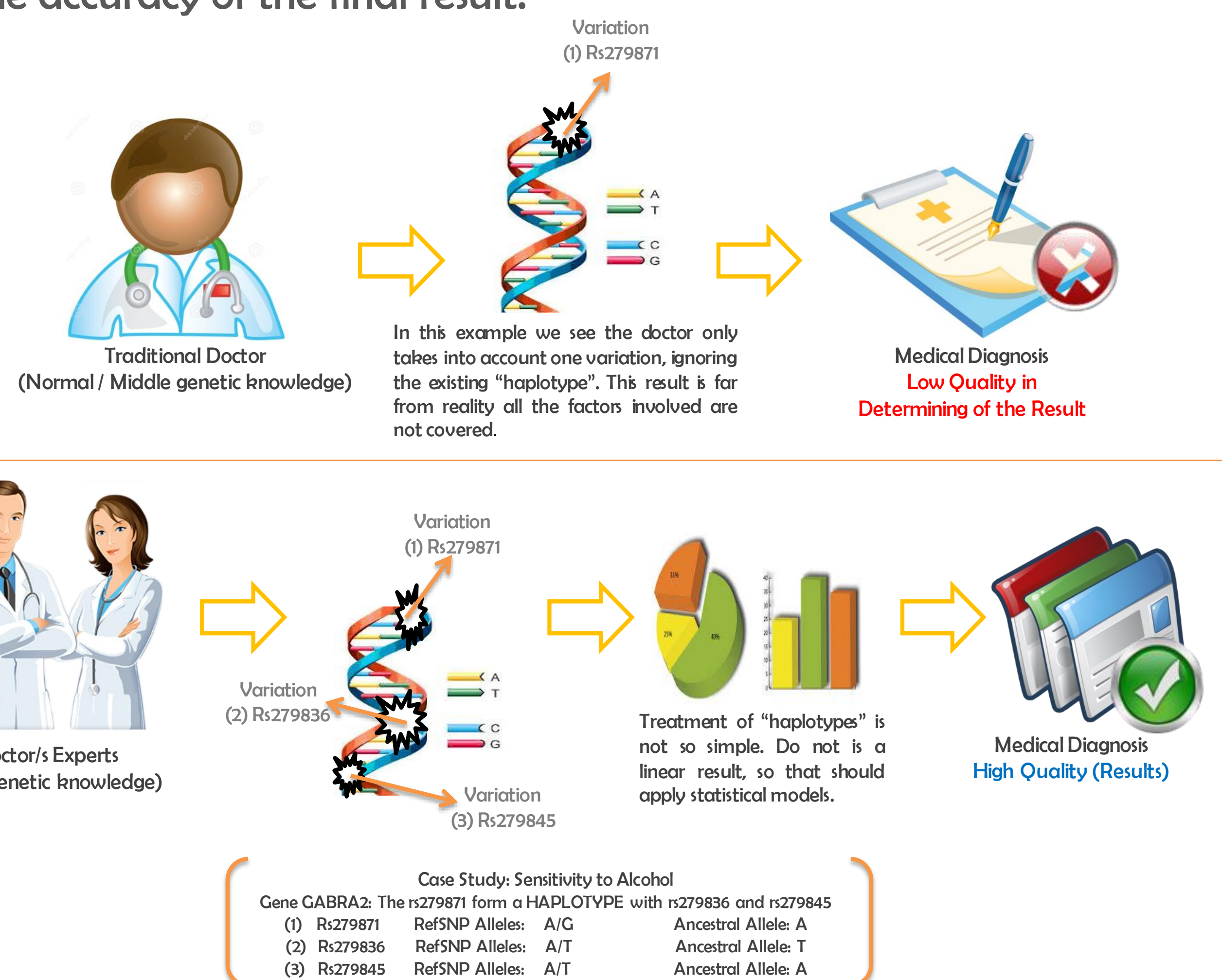
- ✓ Integrating Treatment of haplotypes in CMHG
- ✓ Application of Statistical Models in CMHG



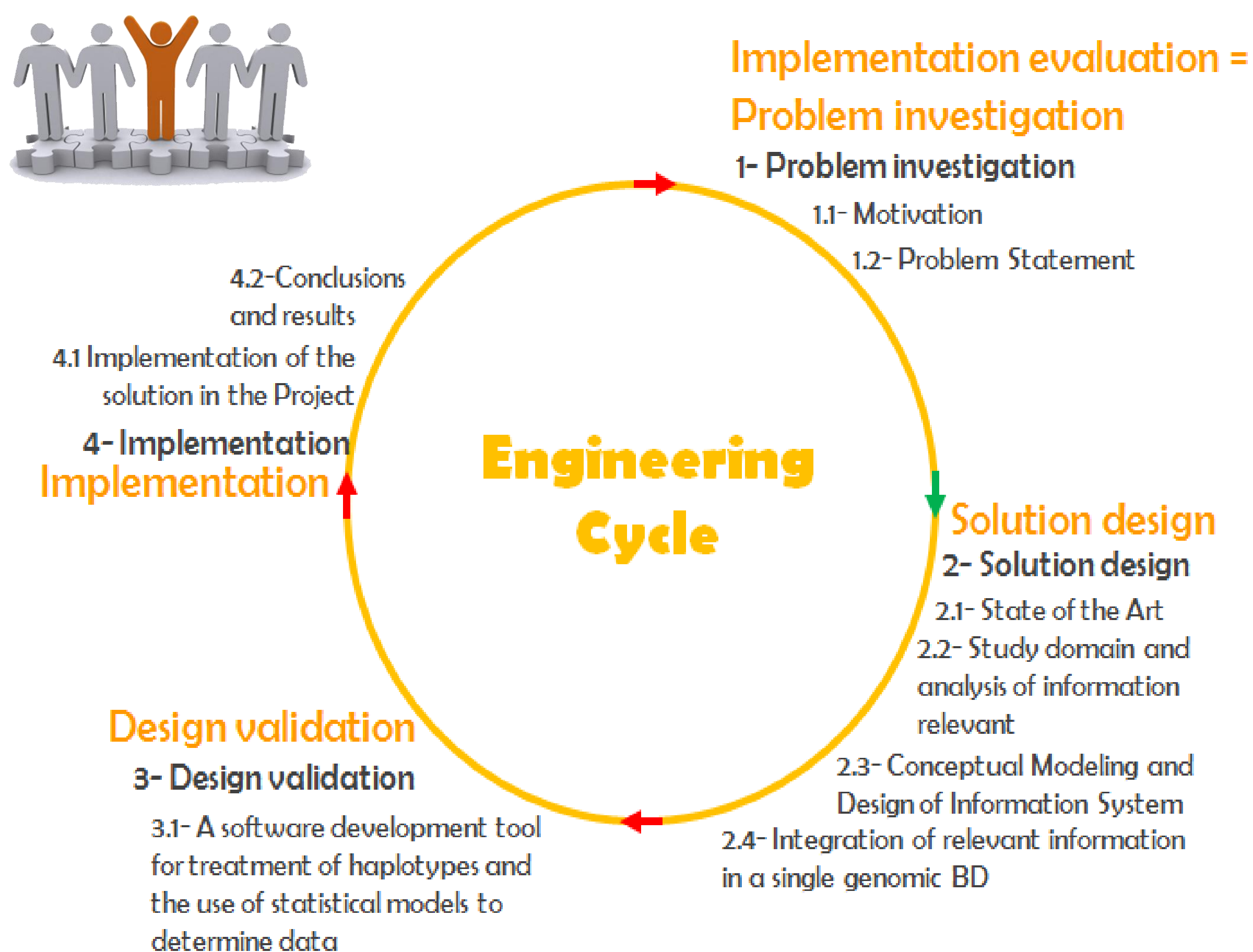
## Results

In this work we start from the Conceptual Model of the Human Genome which does not include the treatment of haplotypes (where variations may intervene or arise from the combination of two or more alleles), so the reality is not so simple. We seek and to include in turn improve the ability to model the generation of Genomic results.

Another contribution is applied on the data quality, the results are not linear, here comes the management of haplotypes causing them to be statistical. The incorporation of statistical models gives giant advances in the accuracy of the final result.



## Research-Phases



The phases identified in this research have been those defined by Roel Wieringa, composed Cycle Engineering, which evolves as follows: 1 Problem Investigation; 2. Solution Design; . 3 Design Validation; 4. Implementation.

## Conclusion



Bioinformatic environment has made great strides in recent years, one of these is the application of Model Driven Development (MDD), which allows us to generate Genome Information Systems (GIS) with improvements to the processing and use of data. In this domain the incorporation of statistical models is an added value that contributes to a determination of data accurately and efficiently.

These models have identified guidelines for the development of a framework that integrates powerful haplotypes and tackling dispersion genomic world. What was impossible a few years back (knowing our genome), is currently a simple and fast (genomic data management) task.