

# A modern Egyptian: the first comprehensive Egyptian reference genome

In this article, we introduce you to the pioneering work of Professor Hauke Busch and Dr Inken Wohlers. Together with their team at the University of Lübeck – and with the help of Novogene *de novo* sequencing and our bioinformatics expertise – Prof. Busch and Dr Wohlers successfully assembled and annotated the world's first comprehensive Egyptian reference genome. It is hoped that this critical work will pave the way for precision medicine throughout North Africa.



## A modern Egyptian: the first comprehensive Egyptian reference genome

The true value of genomics can only be unravelled with true population-scale sequencing. Understanding genomic variation across populations will help propel genetic medicine, providing insight into ancestry, ecology and equity. At a recent Frontline Genomics webinar, Professor Dr Hauke Busch and Dr Inken Wohlers from the University of Lübeck shared their team's work assembling and annotating the world's first comprehensive North African (Egyptian) reference genome (*EgyptRef*). In this project, like many other *de novo* projects worldwide, Novogene were a true collaborative partner, providing the bioinformatic pipelines and software solutions which helped the team maximise production.



## Underrepresented populations

The completion of the Human Genome Project in 2003 revolutionised the genetic field. This international effort took over a decade and cost billions of dollars. Now, with ongoing technological advancements, the ability to sequence an individual's entire genome can be done in a relatively short period of time and at a fraction of the cost. This genetic reference data is essential for precision medicine, as Prof. Busch emphasised.

To date, the majority of genetic studies are based on individuals from European populations. In 2016, it was estimated that 81% of samples from the GWAS Catalog came from individuals of European descent. As a result, many projects throughout the world, including projects in Asia, Uganda and Denmark, have started to actively identify world-wide variation. Nevertheless, Prof. Busch highlighted that there is still very little genetic data available for many regions of the world. In particular, North African individuals are often underrepresented in current datasets, such as the 1000 Genomes and gnomAD databases. Consequently, this has raised concerns about whether information drawn from these studies, particularly GWAS, can be transferable to other populations. These disparities are a major public health concern.

## The first Egyptian genome reference

The aim of Prof. Busch and Dr Wohlers's project, **published** in Nature Communications, was to deliver the first North African genome reference using the latest technological advances. In turn, they hoped this would be the first step towards precision medicine within this region. In collaboration with the University of Mansoura and Novogene, the team from the University of Lübeck generated a *de novo* genome assembly from one male Egyptian individual and used an additional 110 Egyptian individuals to identify single nucleotide (SNV) and structural variants. This data was subsequently integrated to generate a complete Egyptian genome reference.

## The technical detail

The genome was assembled from PacBio, 10x Genomics and Illumina paired-end sequencing data. The team first created two draft assemblies of the personal genome – FALCON (long-read assembly) and WTDBG2 (novel assembly) – which were subsequently polished. The final Egyptian meta-assembly denoted as EGYPT was compared with publicly available assemblies (Korean and Yoruba). The meta-assembly was then complemented with variants from 110 Egyptians. The team called ~20 million SNVs and ~120,000 structural variants. In total, they identified 1,198 population specific SNVs that were rare in public populations. Out of these, the team found that 49 were detected for the first time (did not have a dbSNP ID) and four had deleterious CADD scores.

As most GWAS are based on European populations, inferring variant effect on another population is difficult. In fact, the team found that 261 tagged single nucleotide polymorphisms (7%) were not present in the Egyptian cohort, emphasising the need to perform GWAS in non-European populations. Importantly, they also discovered differences in linkage disequilibrium between Egyptians and Europeans which may compromise GWAS transferability.

Other population genetic analyses conducted involved tracing maternal lineages using mitochondrial haplogroups and undertaking a genotype principal component analysis using public variant data from 143 other populations. From these analyses, the team identified that Egyptians share haplogroups with Europeans (>60%), Africans (24.8%) and Asians (6.7%). They found that the Egyptian cohort were a relatively homogenous group. Admixture analysis

also revealed that the genetics of Egyptian individuals comprised of four distinct population components (75%) – Middle Eastern, European/Eurasian, North African and East African. These results support the notion that Egypt's transcontinental geographical location shaped Egyptian genetics.

### A true collaboration

At Novogene we are proud of the part we have played in this revolutionary project, and excited at the team's monumental findings. It is known that a good high-quality reference is critical for downstream analysis. While short-read data has provided a good foundation, improvements in long-read sequencing technologies have advanced research. Nonetheless, combining and interpreting these different reads into the same genome can propose a challenge. This is where the strengths of Novogene really come to the fore – our ability to not only provide the high quality data required but the high calibre bioinformatic pipelines and software solutions that go with it to help the research teams interpret and maximise production.

This project also fits well with Novogene's purpose to help research and development organisations unlock their ideas and discover solutions that will ultimately improve and extend lives. Dr Yuanyuan Chen, Novogene European Sales Director, commented:

**"We are really excited and honoured to have participated in such cutting-edge scientific discoveries and this is also Novogene's goal – to serve research communities with modern technologies."**

### Just the beginning

This two-year project has not only created the first personal Egyptian reference but has catalysed the study of population genetics within this region. The team have constructed EgyptRef as a community resource. They have published the raw and variant data on the [European Genome-phenome Archive](#) and all summary results and figures on their [website](#).

The team hope that this is now the first step towards precision medicine in North Africa. Prof. Busch reported that the [Egyptian Academy of Scientific Research and Technology](#) has sparked a national genome initiative calling for projects on the Egyptian genome. While this publication just focused on the modern Egyptian, this data could also be compared with data from Ancient Egyptian remains. Prof. Busch stated:

**"It is actually very nice to see this progress in genome research in Egypt now. I think there are a lot of interesting questions one can now ask through this data."**

Not able to attend the webinar on the day? Frontline Genomics has made the full recording available on demand [here](#).

