scientific reports

Check for updates

OPEN

Whole-exome sequencing of the mummified remains of Cangrande della Scala (1291–1329 CE) indicates the first known case of late-onset Pompe disease

Barbara Iadarola[1,5], Denise Lavezzari[1,5], Alessandra Modi[2], Chiara Degli Esposti[1], Cristina Beltrami[1], Marzia Rossato[1], Valentina Zaro[2], Ettore Napione[3], Leonardo Latella[4], Martina Lari[2], David Caramelli[2]✉, Alessandro Salviati[1] & Massimo Delledonne[1]✉

Mummified remains of relevant historical figures are nowadays an important source of information to retrace data concerning their private life and health, especially when historical archives are not available. Next-generation-sequencing was proved to be a valuable tool to unravel the characteristics of these individuals through their genetic heritage. Using the strictest criteria currently available for the validation of ancient DNA sequences, whole-genome and whole-exome sequencing were generated from the mummy remains of an Italian nobleman died almost 700 years ago, Cangrande della Scala. While its genome sequencing could not yield sufficient coverage for in depth investigation, exome sequencing could overcome the limitations of this approach to achieve significantly high coverage on coding regions, thus allowing to perform the first extensive exome analysis of a mummy genome. Similar to a standard "clinical exome analysis" conducted on modern DNA, an in-depth variant annotation, high-quality filtering and interpretation was performed, leading to the identification of a genotype associated with late-onset Pompe disease (glycogen storage disease type II). This genetic diagnosis was concordant with the limited clinical history available for Cangrande della Scala, who likely represents the earliest known case of this autosomal recessive metabolic disorder.

DNA is naturally broken down into fragments after death and is ultimately degraded to single nucleotides, but sequence information can be recovered from samples that are hundreds of thousands of years old[1]. If well-preserved biological samples are available, next-generation sequencing technologies can provide information about historical figures from the recent past, helping to clarify aspects of their official and private lives that cannot be resolved using traditional historical sources. DNA analysis is a non-traditional source of historical information, but it facilities objective and accurate historical reconstructions and is therefore an important resource that can be used to support evidence from traditional sources such as documents, literature and artwork.

An interesting case study is Cangrande della Scala (1291–1329 CE), lord of Verona from 1311 to 1329, a great military commander and politician who brought neighboring cities under his control to form a "kingdom" of the Venetian hinterland spanning from Verona to Treviso. Verifiable data about his private life and health are scarce because the Scaliger family archives were destroyed, forcing historians to rely on less reliable sources that may be influenced by positive or negative bias.

[1]Department of Biotechnology, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy. [2]Department of Biology, University of Florence, Via del Proconsolo 12, 50122 Florence, Italy. [3]UNESCO Office, Municipality of Verona, Piazza Bra 1, 37121 Verona, Italy. [4]Department of Zoology, Natural History Museum of Verona, Lungadige Porta Vittoria 9, 37129 Verona, Italy. [5]These authors contributed equally: Barbara Iadarola and Denise Lavezzari. ✉email: david.caramelli@unifi.it; massimo.delledonne@univr.it

Sample ID	No. of merged reads	GC content (%)	No. of mapped reads before removing PCR duplicates	No. of mapped reads after removing PCR duplicates	Human DNA (%)	Mean coverage (×)
Cangrande liver	4,445,677	34.55	2,572	494	0.06	0.00
Cangrande bone no UDG	7,924,124	41.85	1,856,551	1,533,864	23.43	0.0392
Cangrande bone	6,384,132	41.66	1,488,301	1,099,139	23.31	0.0289

Table 1. Exploratory WGS analysis, indicating the human sequence content in three DNA samples. The table shows the number of sequenced reads, the GC percentage, the number of mapped reads before and after removing PCR duplicates, the percentage of fragments mapped to the human reference genome, and the average coverage of the genome.

Cangrande della Scala was interred in a marble tomb that promoted mummification. His remains were exhumed in 2004 for scientific analysis by a multidisciplinary team of researchers, revealing the presence of digitalis in his well-preserved organs[2,3]. This led to several hypotheses, including murder by poisoning[4] and the therapeutic use of digitalis to remedy a cardiac disorder[4–6]. Here we used samples of bone tissue from the mummified remains for the extraction of ancient DNA, followed by clinical whole-genome sequencing (WGS) and whole-exome sequencing (WES). We identified two pathogenic variants in the *GAA* gene encoding α-glucosidase, a genotype associated with late-onset Pompe disease (also known as acid maltase deficiency, acid α-glucosidase deficiency, and glycogen storage disease type II). The clinical phenotype of this disease is consistent with data from the historical records, suggesting that Cangrande della Scala is the earliest known case of this prototypic lysosomal storage disorder.

Results

WGS performance and authentication of ancient DNA. DNA extracted from the mummified remains of Cangrande della Scala (right intermediate cuneiform bone and liver) was used for three exploratory WGS experiments. Two WGS datasets were prepared from bone DNA, one including partial uracil-DNA glycosylase treatment and one with no treatment. A third dataset was prepared from liver DNA to assess the degree of DNA preservation. The percentage of human sequence content in all three samples was compared by alignment to the human reference genome (Table 1). This revealed an extremely low percentage of mapped fragments from the liver sample (0.06%) but a much higher percentage from the bone sample (23%).

The fraction of human DNA recovered from the cuneiform bone sample was consistent with values obtained from other small foot bones (phalanxes) in human remains of similar age[7]. We carried out several tests to confirm the ancient nature of the sequenced human DNA and to exclude modern DNA contamination. The tests revealed that the library sequences showed features typical of degraded DNA, such as an average length of ~ 90 bp without uracil-DNA glycosylase treatment and 86 bp with partial uracil-DNA glycosylase treatment. Furthermore, the frequency of 5′ cytosine deamination in the library prepared without uracil-DNA glycosylase treatment was consistent with the age of the sample[8] (Supplementary Table S1, Supplementary Figs. S1 and S2). Complete mitochondrial genomes could be reconstructed from both libraries (mean coverage of 11.42 × and 8.22 × respectively) and both sequences were unambiguously assigned to the same mitochondrial haplogroup (X2b11) with a maximum score (Supplementary Table S2). Finally, present-day human DNA contamination in mitochondrial sequences, estimated using two different methods, did not exceed 1% (Supplementary Table S2). Genetic sex determination revealed Ry values of 0.0885 and 0.0887 for the libraries prepared with and without uracil-DNA glycosylase treatment, respectively, confirming that the bone DNA belonged to a male individual (Supplementary Table S3).

Exome capture and WES performance. Having confirmed satisfactory DNA preservation, we used the bone sample for WES to enrich for human sequences and achieve good coverage for variant calling at a reasonable cost. To limit the effect of DNA damage on variant calling, whole-exome capture was restricted to the library prepared using partial uracil-DNA glycosylase treatment. Quality control of the enriched library revealed a size range of 116–769 bp (average = 277 bp) and a concentration of 4.68 ng/µl (Fig. 1).

The WES library was sequenced twice, yielding 62 million and 77 million fragments, respectively (Table 2). The mapped coverage on the exome target was on average 29.35 × in the first run, with a large proportion of PCR duplicates (51.72%). This increased slightly to 31.83 × in the second run, but again the low coverage reflected the large proportion of PCR duplicates (55%). When the two runs were combined, the mapped coverage was 37.28 × and the percentage of duplicates increased to 65%, indicating that library saturation had been achieved. Despite the low average coverage, the high uniformity of enrichment (FOLD80 penalty ≥ 1.41) allowed us to cover 99.16% of the exome target with at least 10 reads (Twist design) and to genotype 93.38% of the target bases (Table 2).

Exome data authentication and estimate of nuclear DNA contamination. The misincorporation patterns and fragment sizes of the WES data (Supplementary Fig. S3) were consistent with the profiles derived from the WGS data (Supplementary Fig. S2): 2.17% C > T transitions at the first 5' base and an average fragment length of 100 bp, approximately 10 bp longer than the pre-capture size, in agreement with earlier studies[9–11] (Supplementary Table S4). The WES data were also used to estimate putative nuclear DNA contamination. Taking advantage of the male genetic sex of the sample, we measured the heterozygosity observed at 1344 polymor-

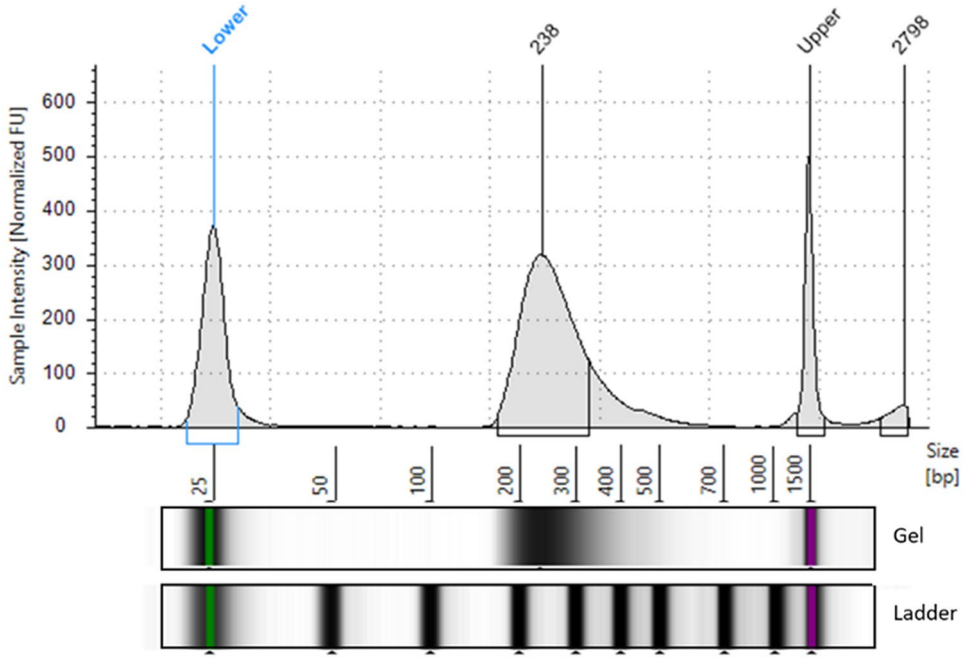

Figure 1. Distribution of fragment sizes in the enriched library used for WES.

	No. of fragments	No. of mapped reads before removing PCR duplicates	Percent dupl	No. of mapped reads after removing PCR duplicates	Mean coverage (×)	%10×	%PASS	FOLD80
RUN1	62,394,319	107,031,260	51.72	51,669,825	29.35	98.08	93.66	1.53
RUN2	77,921,888	135,479,585	55.99	59,615,288	31.83	98.60	93.60	1.39
MERGED RUNS	140,316,207	242,510,843	65.47	83,738,060	37.28	99.16	93.38	1.41

Table 2. Performance of the two WES runs. The table shows the number of sequenced fragments, the number of mapped reads before and after duplicate removal, the percentage of duplicated reads, the average coverage of the target, the percentage of the target covered at least by 10 reads, the genotypability, and the uniformity of coverage (FOLD80 penalty value).

phic sites on the X-chromosome. This revealed 0.0047% X-chromosome contamination, with an estimated error of 2.177062×10^{-3} thus confirming the authenticity of the ancient DNA data (Supplementary Table S4).

Variant identification and prioritization. We identified 24,769 variants in the exome target regions, 34.15% of which were homozygous (33% SNVs and 1% INDELs) and 65.85% of which were heterozygous (63% SNVs and 2% INDELs). The overall Ti/Tv ratio was 2.77, which is consistent with other WES studies[12]. The variants were then investigated to determine whether Cangrande della Scala carried mutations that may have contributed to his death. Variants were prioritized using two dedicated pipelines (annotated variants and predicted coding variants, as described in the methods section). Among the final set of prioritized variants (249 annotated variants and 1342 predicted coding variants, Supplementary Tables S5 and S6), those classified in HGMD and/or ClinVar were investigated in detail. This reduced the priority list to 210 ClinVar variants and 179 HGMD variants, 140 of which were common to both databases. Subsequent analysis focused on 113 rare clinical variants associated with more severe diseases. They included two heterozygous missense mutations in two exons of the *GAA* gene associated with the autosomal recessive phenotype "late-onset Pompe disease". The first variant (c.1465G>A) was classified as "pathogenic/likely pathogenic" in ClinVar and damaging (DM) in HGMD, whereas the second (c.271G>A) was classified as "benign/likely benign" in ClinVar and likely damaging (DM?) in HGMD. Moreover, we identified additional variants in genes correlating with the regulatory activity of lysosomal enzymes[13]: besides 28 variants which were common in the frequency population databases (MAF > 5%), six rare exonic missense variants were present in the *ATP6* gene and one rare exonic missense variant was present in the *RUNX1* gene.

Phasing the *GAA* variants. Chromosome-wide phasing was used to determine the cis/trans phase of the two heterozygous genotypes in the *GAA* gene, considering all the variants on chromosome 17. Only one of the two variants of interest (c.271G>A) could be correctly phased, whereas the other (c.1465G>A) was discarded by

the algorithm because it was missing from the reference population. To exclude the possibility that the two variants did not constitute a haplotype, we assessed the longest haplotype in the reference population carrying variant c.271G>A. The aim was to confirm that the haplotype containing c.271G>A spanned over the c.1465G>A variant position, which could be inferred as a reference for this haplotype in the population. This allowed the identification of a set of variants consistently inherited together with c.271G>A in 98% of reference individuals, namely in a linkage disequilibrium block more than 14 kb in length (chr17:80,096,549–80,110,889) and spanning over the c.1465G>A variant position. It is therefore very unlikely that the latter variant was inherited with c.271G>A on the same chromosome. Moreover, the frequency of the two variants differed in the general population frequency databases, confirming the low probability of both being inherited together (c.271G>A allele frequency = 3% in all the population databases, c.1465G>A allele frequency reported in two databases only, with values of 0.0009% and 0.0015%, respectively). These data support the hypothesis of a compound heterozygous genotype (variants affecting the *GAA* gene in trans) thus confirming the diagnosis of late-onset Pompe disease.

Discussion

We have described the successful clinical analysis of a 700-year-old human mummy by WES and demonstrated that exome enrichment applied to ancient human DNA can lead to a genetic diagnosis that may help to support historical data. By integrating WGS and WES, we confidently assessed the authenticity of the data and excluded significant bias caused by contamination with modern human DNA.

Next-generation sequencing for the study of ancient samples has mainly targeted small regions, such as the mitochondrial DNA (mtDNA), the Y-chromosome DNA (Y-DNA)[14] or specific SNPs of interest[15,16]. Only a few "high-quality" ancient human WGS studies have been reported, including a Denisovan[17] (30 × mean coverage) and two Neanderthal individuals (52 × and 27 × mean coverage)[18,19] investigated mainly to discover gene flow events and admixtures of archaic hominins. Typically, low-coverage genomes are used to investigate genetic diversity at the population level and provide phylogenetic information[20,21]. Considering the low quantity of human material present in Cangrande's samples, which would have required an abnormal sequencing effort, we considered WES as a much more affordable technology to eliminate environmental DNA contaminants and enrich for protein-coding regions, which are mainly responsible for the development of Mendelian disorders[22]. WES has already been applied in a limited number of studies, although these did not focus on the functional consequences of genetic variants and their association with clinical phenotypes[23,24].

WES performed on samples from the mummified remains produced 37 × coverage of the protein-coding regions. The genotypability values and number of identified variants were almost comparable to the WES analysis of modern humans[25]. We identified two clinically relevant compound heterozygous variants in the *GAA* gene associated with late-onset Pompe disease, which could very well explain Cangrande's disease and death.

Contemporaneous sources on Cangrande's life provide little verifiable information about his health. These documents are also difficult to interpret because the authors were either very close to the court or enemies of the Scaligeri. Most accounts of Cangrande's childhood are laudatory and are based on literary *topoi*. They describe a child who did not like traditional games or the company of peers, but was predisposed to military life. We know from the pro-Scaliger writer Ferreto Ferreti and from the Paduan enemy Albertino Mussato that Cangrande, even in combat on horseback, preferred to use the bow in the Parthian manner rather than the spear or sword, allowing his arm more freedom of movement[26,27]. Cangrande was reportedly ill for some time at the age of 23 but recovered enough for battle after imbibing a small dose of an antidote (not better defined) and a sip of wine[28]. The enemy chronicler Mussato attributed the discomfort to one foot, which prevented the Scaligero from riding[27]. He was forced to abandon his horse and accept the draft horse of a peasant[28]. At the age of 29, Cangrande was reportedly pierced by an arrow in the thigh[29], but managed to return to camp, rally his troops and return to fight[30]. The autopsy of Cangrande in 2004 did not reveal any wounded limbs, suggesting that Cangrande instead suffered crippling thigh discomfort, possibly a cramp. In his next battle he once again had to abandon his horse and accept one from a peasant[31]. At the age of 34, Cangrande fell seriously ill for a long time and was given up for dead[32]. Cangrande died at the age of 38, after showing symptoms of malaise for 3 days identified as fever and a generic fluxum, which can be translated as vomiting or perhaps hemorrhage[31]. Some sources report fluxus ventris or intestinal disease with diarrhea, which was disproven in the 2004 autopsy by the discovery and examination of solid fecal matter at the base of the rectum. Cangrande retained his mental clarity at the end of his life, enabling the completion of an important juridical document[32].

The clinical spectrum (infantile, juvenile and adult-onset) of Pompe disease is a continuum, depending on the residual activity of α-glucosidase. In late-onset forms, the storage of glycogen is confined to the skeletal muscle, heart and liver. The clinical manifestation includes skeletal muscle weakness, respiratory distress due to diaphragm and accessory respiratory muscle weakness, muscle cramps, spontaneous bone fractures and arrhythmogenic cardiopathy (but normal intelligence). This is entirely consistent with the Cangrande's three episodes of severe weakness after exercise as reported in the historical records, and his death after 3 days of sickness, but with no mental impairment. No scar was found on his thighs, supporting the hypothesis that the arrow wound reported in 1320 was in fact a severe muscle cramp. Given the finding of digitalis in his body[4], it is possible that it was administered to counteract tachycardia, a key symptom of cardio-respiratory insufficiency, and would represent the first known clinical use of this drug.

The first variant (c.1465G>A) found in Cangrande's *GAA* gene is described in patients with late-onset Pompe disease and fully inhibits α-glucosidase maturation and activity[33]. The second variant (c.271G>A), known as *GAA* *2 and reported as likely damaging in the human database of HGMD, is frequent in the Caucasian English population (0.03)[34]. The biochemical phenotype associated to this genotype shows reduced activity toward the natural substrate (glycogen) but normal activity toward the artificial substrate 4-methylumbelliferil-α-glucopyranoside. The α-glucosidase encoded by the *GAA**2 allele has a K_m for glycogen tenfold higher than the wild-type enzyme

and enzymatic activity towards glycogen that is 1/10 of normal[34]. To explain the absence of this frequent allele in their 15 late-onset cases, Swallow et al.[34] cited a modification of the Michaelis–Menten equation[35] where the increase in K_m is offset by an increase in substrate concentration and therefore they did not classify variant *GAA* *2 as a disease-causing allele, based on the assumption that lysosomal pathology is due to encumbrance. Since in our case such variant was instead present in trans-configuration with a pathogenic allele producing no enzyme activity (c.1465G>A), we concluded that Cangrande had a α-glucosidase with just 10% of normal activity. While abrogation of α-glucosidase enzymatic activity is causative of the classical infantile form, such condition is instead responsible of the late-onset Pompe disease, associated with α-glucosidase activity lower than 20%[36,37]. Additional genes have been recently demonstrated to play a role as genetic modifiers of lysosomal functions[13,38,39]. Among these we found that Cangrande genome carried six rare missense variants in *ATP6* (controlling the pH of the lysosomal compartment)[38,39], while *RUNX1* (involved in autophagy/lysosome metabolism)[13] had one rare missense variant. Beside the *GAA* alleles causative of the late-onset Pompe disease, these other genetic variants could also contribute to Cangrande's clinical phenotype.

Methods

Conservation of the biological remains. In 2007, the biological remains of Cangrande della Scala that had not been reinterred were deposited in the Natural History Museum of Verona to be preserved and made available for further analysis. The remains (parts of the liver, phalanges, metatarsal and cuneiform bones) were placed in sterile receptacles and stored in the dark at 19–21 °C and 40–45% relative humidity, with periodic monitoring. These are standard conditions for the storage of biological material[40]. The selection and collection of samples was carried out in the Zoology Laboratory of the Natural History Museum of Verona.

Sample preparation and DNA extraction. DNA isolation and library preparation were carried out at the Molecular Anthropology and Paleogenetic Laboratory of the Department of Biology, University of Florence, using facilities exclusively dedicated to ancient DNA analysis, following stringent protocols to prevent contamination with present-day DNA[41]. Negative controls were included in each experimental step. The right intermediate cuneiform bone and a small portion of mummified liver tissue were collected for DNA analysis. To remove potential contaminants, the outer layer of the bone sample was brushed with disposable tools and irradiated with ultraviolet light (254 nm) for 45 min in a Biolink DNA Crosslinker (Biometra). Bone powder was then collected from the densest part of the bone using a low-speed dental micromotor equipped with disposable tungsten carbide ball burrs. DNA was extracted from 50 mg of bone powder using a silica-based protocol that allows DNA molecules to be recovered efficiently even if highly fragmented[9]. In the final step, DNA was eluted twice in 50 µl TET buffer (10 nM Tris, 1 mM EDTA, 0.05% Tween-20). DNA was extracted from 50 mg of mummified liver tissue using the QIAamp DNA mini kit according to the manufacturer's recommendations (Qiagen).

Whole-genome library preparation and sequencing. Sequencing libraries suitable for Illumina platforms were prepared from 20 µl of DNA extracted from bone or liver tissue following a protocol optimized for ancient samples[42]. The resulting data were used to evaluate the deamination rate to confirm the authenticity of the genetic material. A partial uracil-DNA glycosylase treatment[43] was applied to an additional 30-µl aliquot of DNA extracted from bone. This treatment removes internal uracil residues and abasic sites, reducing the probability of errors during variant calling. A unique combination of two indices per library was used for barcoding. Libraries were sequenced in 150-bp paired-end mode on a NovaSeq 6000 instrument (Illumina) to generate an average 1 × coverage of the entire genome.

Bioinformatics analysis of WGS data: preservation and contamination estimates, molecular sex determination, and mitochondrial genome reconstruction. Sequences were demultiplexed and sorted according to the indices, and raw sequence data from all three libraries were analyzed using an established pipeline[44]. Adapters were clipped-off and reads with a minimum overlap of 10 bp were merged in a single sequence using Clip&Merge v1.7.4. Merged reads were then mapped onto GRCh38 using BWA v0.7.17-r1188[45], setting parameters to improve the accuracy of ancient DNA reads (-l = 16,500, -o = 2 and -n = 0.01)[46]. Only reads with a map and base qualities score ≥ 30 were retained. Reads mapped onto the human genome were authenticated by deamination and fragmentation pattern analysis using mapDamage2.0[47].

Molecular sex determination was applied to the bone sample by comparing the number of alignments to the Y chromosome and the total number of alignments to the X and Y chromosomes in the libraries prepared with and without uracil-DNA glycosylase[48]. The mitochondrial genome was reconstructed from the same libraries to assess overall DNA preservation and contamination. Reads mapping to the mitochondrial genome were extracted from BAM files using SAMtools v1.7[49]. For the library prepared without uracil-DNA glycosylase, the Schmutzi pipeline[50] was used to call the consensus sequence and to evaluate the level of contamination with present-day human DNA[51]. For the library prepared using a partial uracil-DNA glycosylase treatment, the consensus sequence was called using mpileup and vcfutils.pl in the SAMtools package. To estimate the ratio of contaminant/authentic DNA in the mitochondrial sequence data, a likelihood-based method was used as previously described[52]. The mitochondrial haplogroups were assigned according to PhyloTree build 17[53] using Haplogrep2[54].

Exome enrichment and WES. The DNA library prepared from bone was captured using the Twist Bioscience Human Core Exome Kit + RefSeq v1.3 protocol. Single-plex exome capture was carried out using 120-bp biotinylated probes, with minor modifications to the standard protocol due to the high level of sample degradation. All the available material from the library was used, although the DNA input requirements of the standard protocol were higher than the sample's initial input (300 ng instead of 500 ng). After the washing steps to remove

nonspecific targets, the remaining material was eluted in 22.5 µl of water, without keeping the backup slurry. Ten cycles of amplification were performed rather than the eight cycles recommended by the protocol. The final PCR cleanup was carried out using a 1.5 × ratio of Twist Bioscience Beads. The enriched library was validated using a Tape Station 4150 High Sensitivity D1000 assay kit (Agilent Technologies) and quantified by RT-PCR using the Lib Quant kit (Roche). WES was then performed on a NovaSeq 6000 instrument in 2 × 100-bp paired-end mode.

Bioinformatics analysis on WES data: read alignment, variant calling, and data authentication. The WES FASTQ files were quality checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapters and low-quality bases were removed, and reads were aligned with the human reference genome (GRCh38/hg38) using the Paleomix bam_pipeline v1.2.13.8[55] with BWA-mem v0.7.17[45] and a disabled "–collapse" parameter to properly calculate the insert size for all the sequenced fragments. Duplicated reads were removed using Picard MarkDuplicates v2.21.1. GATK Base Recalibrator v4.1.8.1[56] and BamUtil clipoverlaps v1.4.14 were then applied to adjust base quality and soft-clip overlapping reads. Coverage and genotypability metrics were calculated using CallableLoci in GATK v3.8. The FOLD80 penalty value was calculated using Picard CollectHSMetrics v2.21.1 (http://broadinstitute.github.io/picard/). Variants were identified using GATK HaplotypeCaller v4.1.8.1 (with parameter "–dont-use-soft-clipped-bases" set to "true"), producing a gVCF file. Variants were then recalibrated and filtered by following the GATK Hard Filtering Best Practices.

The authenticity of the WES data was estimated as previously described for WGS. Additionally, nuclear contamination was estimated by measuring the heterozygosity of the X chromosome[57] using the ANGSD pipeline[58]. Because males have only one copy of the X chromosome, any heterozygosity on this chromosome in males indicates contamination.

Variant annotation, prioritization and phasing. The gVCF file was annotated using Golden Helix VarSeq v.2.2.1 (Golden Helix) and the following databases: ClinVar and HGMD Professional v2020.1[59] were used to investigate the clinical significance of identified variants, whereas the population frequency databases of 1000Genomes Project Phase3, gnomAD v2.0.1 and ESP6500 v2 were used to determine the frequency of variants. Similarly, an internal database was used to flag rare Italian variants. Several prediction tools were used to calculate the pathogenicity scores for each genetic variation (FATHMM, GERP, Polyphen, SIFT, PhastCons and PhyloP) and the RefSeq Genes database was integrated to provide the effect of each variant. Variants were then prioritized using two pipelines. The "annotated variants" pipeline retained only those variants classified as "Pathogenic", "Likely Pathogenic", "Conflicting", "Uncertain Significance" or "Other" in ClinVar, or classified as "DM" or "DM?" in HGMD. Variants with an alternative allele frequency below 5% in the population frequency databases were flagged. The "predicted coding variants" pipeline retained variants present in exonic or splicing site regions but without reported clinical significance in ClinVar and HGMD. Variants with an alternative allele frequency below 1% in the population frequency databases were flagged, focusing on those with: (1) a "LOF", "Missense" or "Splice_region_variants" effect in the RefSeq database; (2) a predicted "Damaging" effect by three or more prediction tools applied to the dbNSFP database (SIFT, Polyphen2, MutationTaster, MutationAssessor and FATHMM); and (3) a frequency below 2% in the Functional Genomics Variant database. Phasing of WES reads was accomplished using Eagle v2.4.1, with the provided hg38 genetic map and the reference panel from gnomAD v3.1 "HGDP + 1 KG callset" (https://gnomad.broadinstitute.org/downloads#v3-hgdp-1kg).

Data availability

The WES variants data are available for download from our public repository using the link: http://ddlab.sci.univr.it/files/Cangrande/Cangrande.tar.gz (VCF file with associated BED files of callable regions).

Received: 8 June 2021; Accepted: 14 October 2021
Published online: 26 October 2021

References

1. Allentoft, M. E. *et al.* The half-life of DNA in bone: Measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B Biol. Sci.* **279**, 4724–4733 (2012).
2. Atoche, P. & Aufderheide, A. C. *Mummies and Science World Mummies Research VI World Congress* (Springer, 2008).
3. Fornaciari, G. *et al.* A medieval case of digitalis poisoning: The sudden death of Cangrande della Scala, lord of verona (1291–1329). *J. Archaeol. Sci.* **54**, 162–167 (2015).
4. Napione, E. *Il Corpo Del Principe: Ricerche su Cangrande della Scala* (Springer, 2006).
5. Perciaccante, A. *et al.* Lessons from the past: Some histories of alpha-1 antitrypsin deficiency before its discovery. *COPD J. Chronic Obstr. Pulm. Dis.* **15**, 1–3 (2018).
6. Schulz, J. *La morte di Cangrande I della Scala: Un caso da riaprire* (Springer, 2015).
7. Parker, C. *et al.* A systematic investigation of human DNA preservation in medieval skeletons. *Sci. Rep.* **10**, 18225 (2020).
8. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7**, e34131 (2012).
9. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* **110**, 15758–15763 (2013).
10. Ozga, A. T. *et al.* Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. *Am. J. Phys. Anthropol.* **160**, 220–228 (2016).
11. Modi, A. *et al.* Combined methodologies for gaining much information from ancient dental calculus: Testing experimental strategies for simultaneously analysing DNA and food residues. *Archaeol. Anthropol. Sci.* **12**, 1–11 (2020).
12. Zhang, W. Q. *et al.* Comparing genetic variants detected in the 1000 genomes project with SNPs determined by the International HapMap Consortium. *J. Genet.* **94**, 731–740 (2015).
13. Napolitano, F. *et al.* Rare variants in autophagy and non-autophagy genes in late-onset pompe disease: Suggestions of their disease-modifying role in two Italian families. *Int. J. Mol. Sci.* **22**, 3625 (2021).

14. Gaudin, M. & Desnues, C. Hybrid capture-based next generation sequencing and its application to human infectious diseases. *Front. Microbiol.* **9**, 27 (2018).
15. Schuenemann, V. J. *et al.* Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat. Commun.* **8**, 1–11 (2017).
16. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
17. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
18. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
19. Mafessoni, F. *et al.* A high-coverage neandertal genome from chagyrskaya cave. *Proc. Natl. Acad. Sci. USA.* **117**, 15132–15136 (2020).
20. Gravel, S. *et al.* Reconstructing native American migrations from whole-genome and whole-exome data. *PLoS Genet.* **9**, 1004023 (2013).
21. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3**, 1–9 (2012).
22. Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Disease gene identification strategies for exome sequencing. *Eur. J. Hum. Genet.* **20**, 490–497 (2012).
23. Olalde, I. *et al.* Genomic analysis of the blood attributed to Louis XVI (1754–1793), king of France. *Sci. Rep.* **4**, 1–7 (2014).
24. Castellano, S. *et al.* Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci. USA.* **111**, 6666–6671 (2014).
25. Iadarola, B. *et al.* Shedding light on dark genes: Enhanced targeted resequencing by optimizing the combination of enrichment technology and DNA fragment length. *Sci. Rep.* **10**, 1–11 (2020).
26. Cipolla, C. *Le opere di Ferreto de' Ferreti vicentino, III, De Scaligerorum origine poema* (Springer, 1920).
27. Mussato, A. *De gestis Italicorum Post Henricum VII Cesarem (Libri I-VII)* (Springer, 2019).
28. Cipolla, C. *Le opere di Ferreto de' Ferreti Vicentino: Historia Rerum in Italia Gestarum ab Anno MCCL ad Annum Usque MCC-CXVIII.* (1914).
29. Mussato, A. *De Gestis Italicorum Post Henricum VII Cesarem, seu de Conflictu Domini Canis Grandis de Verona Apud Moenia Paduanae Civitatis. Liber XI.* (1727).
30. Pagnin, B. *Guillelmi de Cortusiis Chronica de novitatibus Padue et Lombardie.* (1941).
31. Vaccari, R. *Chronicon Veronense di Paride da Cerea e dei Suoi Continuatori (Il), II/1, La Continuazione Scaligera (1278–1375).* (2014).
32. Varanini, G. M. *La Morte di Cangrande della Scala. Strategie di Comunicazione Intorno al Cadavere, in Cangrande della Scala. La Morte e il Corredo di un Principe nel Medioevo Europeo*. (2004).
33. Montalvo, A. L. E. *et al.* Mutation profile of the GAA gene in 40 Italian patients with late onset glycogen storage disease type II. *Hum. Mutat.* **27**, 999–1006 (2006).
34. Swallow, D. M. *et al.* An investigation of the properties and possible clinical significance of the lysosomal?-glucosidase GAA 2 allele. *Ann. Hum. Genet.* **53**, 177–184 (1989).
35. Conzelmann, E. & Sandhoff, K. Partial enzyme deficiencies: Residual activities and the development of neurological disorders. *Dev. Neurosci.* **6**, 58–71 (1983).
36. Musumeci, O. *et al.* LOPED study: Looking for an early diagnosis in a late-onset Pompe disease high-risk population. *J. Neurol. Neurosurg. Psychiatry* **87**, 5–11 (2016).
37. Mehler, M. & Dimauro, S. Residual acid maltase activity in late-onset acid maltase deficiency. *Neurology* **27**, 178 (1977).
38. Settembre, C., Fraldi, A., Medina, D. L. & Ballabio, A. Signals from the lysosome: A control centre for cellular clearance and energy metabolism. *Nat. Rev. Mol. Cell Biol.* **14**, 283–296 (2013).
39. Ballabio, A. & Bonifacino, J. S. Lysosomes as dynamic regulators of cell and organismal homeostasis. *Nat. Rev. Mol. Cell Biol.* **21**, 101–118 (2020).
40. Samadelli, M., Roselli, G., Fernicola, V. C., Moroder, L. & Zink, A. R. Theoretical aspects of physical-chemical parameters for the correct conservation of mummies on display in museums and preserved in storage rooms. *J. Cult. Herit.* **14**, 480–484 (2013).
41. Llamas, B. *et al.* From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *Sci. Technol. Archaeol. Res.* **3**, 1–14 (2017).
42. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **5**, 5448 (2010).
43. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil DNA glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. B* **370**, 2013624 (2015).
44. Peltzer, A. *et al.* EAGER: Efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
45. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
46. Schubert, M. *et al.* Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
47. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
48. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
49. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
50. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 776 (2015).
51. Modi, A., Vai, S. & Posth, C. More data on ancient human mitogenome variability in Italy: New mitochondrial genome sequences from three Upper Palaeolithic burials. *Submitted*.
52. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
53. van Oven, M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. Int. Genet. Suppl. Ser.* **5**, e392–e394 (2015).
54. Weissensteiner, H. *et al.* HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
55. Schubert, M. *et al.* Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).
56. Auwera, G. A. *et al.* From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1-11.10.33 (2013).
57. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
58. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
59. Stenson, P. D. *et al.* The human gene mutation database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum. Genet.* **139**, 1197–1207 (2020).

Acknowledgements
This research was carried out in the framework of the Joint Project 2018 "The genome of Cangrande della Scala: DNA as historical source".

Author contributions
Conceptualization, M.D., L.L., E.N.; methodology, D.C., M.L. M.D.; resources, L.L.; software, B.I., D.L.; validation, A.S.; formal analysis, B.I., D.L, A.S., A.M, V.Z.; investigation, C.D.E, C.B.; writing—original draft, B.I., D.L, M.D., A.S.; writing—review and editing, E.N., M.L., D.C., A.S., M.R., M.D.; supervision, M.R., M.D., A.S.; funding acquisition, L.L and M.D.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-00559-1.

Correspondence and requests for materials should be addressed to D.C. or M.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.