Genetic and Epigenetic study of Formalin-damaged DNA with duet multiomics solution evoC 6-base sequencing

Introduction

Formalin-fixed, paraffin-embedded (FFPE) specimens are a common source of long-term stored samples used in research or clinical settings across a range of applications including oncology, developmental biology, histopathology, and many others. FFPE samples represent an important way of conserving precious material that enable investigations to be carried out across an extended period of time, during which a fresh sample would experience significant degradation and informational loss. However, some alterations can be observed in samples due to the fixation process, and therefore any techniques that assess FFPE samples must take into account these potential changes.

Studies that incorporate both genetic and epigenetic information such as DNA methylation (both 5mC & 5hmC) add additional understanding into fundamental pathways of gene activation or silencing. This deeper understanding can give actionable insights into conditions or diseases with real world connotations, such as enabling earlier detection of cancer. The ability to derive combined genetic and epigenetic data from FFPE samples can allow researchers to capture meaningful biological insight from precious stored samples that would otherwise be impossible to uncover. However, DNA damage induced by formalin fixation (such as deamination, fragmentation or nucleic acid crosslinking) can lead to decreased genetic and epigenetic data quality when using next generation sequencing (NGS) approaches.

Here we apply a novel 6-base sequencing technology, duet multiomics solution evoC, to DNA extracted from both formalin-compromised DNA (fcDNA), as well as matched FFPE and fresh-frozen cancer samples, to explore how DNA methylation profiles may become altered by the fixation process. This innovative 6-base sequencing approach offers an enzymatic single-workflow solution enabling the simultaneous detection of both cytosine modifications (5mC, 5hmC) alongside canonical bases (A, C, G, T) at single-base resolution. Simultaneous detection minimises the information loss experienced with other methods that require multiple workflows to achieve the same result, while preserving the ability to discriminate important C-to-T transitions. This study will investigate the quality of 6-base genomes derived from FFPE material, and assess the impact of various levels of formalin exposure on the resulting epigenetic sequencing results.

Figure 1: duet evoC provides multimodal data without compromised accuracy. The fully enzymatic workflow provides complete and accurate 6-base genome information and is compatible with FFPE samples. It delivers reference quality SNP calling without ambiguity in C-to-T mutation calls and high sensitivity and specificity whilst distinguishing 5mC from 5hmC.

Method

To evaluate the performance of duet multiomics solution evoC on FFPE samples, sequencing libraries were prepared using fcDNA reference standards obtained from Horizon and a variety of matched FFPE and fresh frozen cancer samples. fcDNA samples mimic the types of DNA damage typically seen in FFPE samples and vary in levels of formalin damage (mild, moderate, severe). FFPE DNA samples were extracted from Colorectal (CRC) and Lung cancer tissues with matched fresh-frozen tissue samples from the same patient.

Results

We used the Agilent TapeStation system and genomic DNA screentape assay to determine the integrity of each DNA sample. A DIN (DNA Integrity Number) value is calculated by the software (ranging from 1-10), with lower DIN values indicating higher levels of DNA degradation.

fcDNA samples with mild, moderate and severe formalin damage levels corresponded with DIN numbers of 6.5, 2.8 and 1.6 respectively. FFPE cancer samples had moderate damage with DIN values of 3.8 and 3.9, shown in Figure 2.

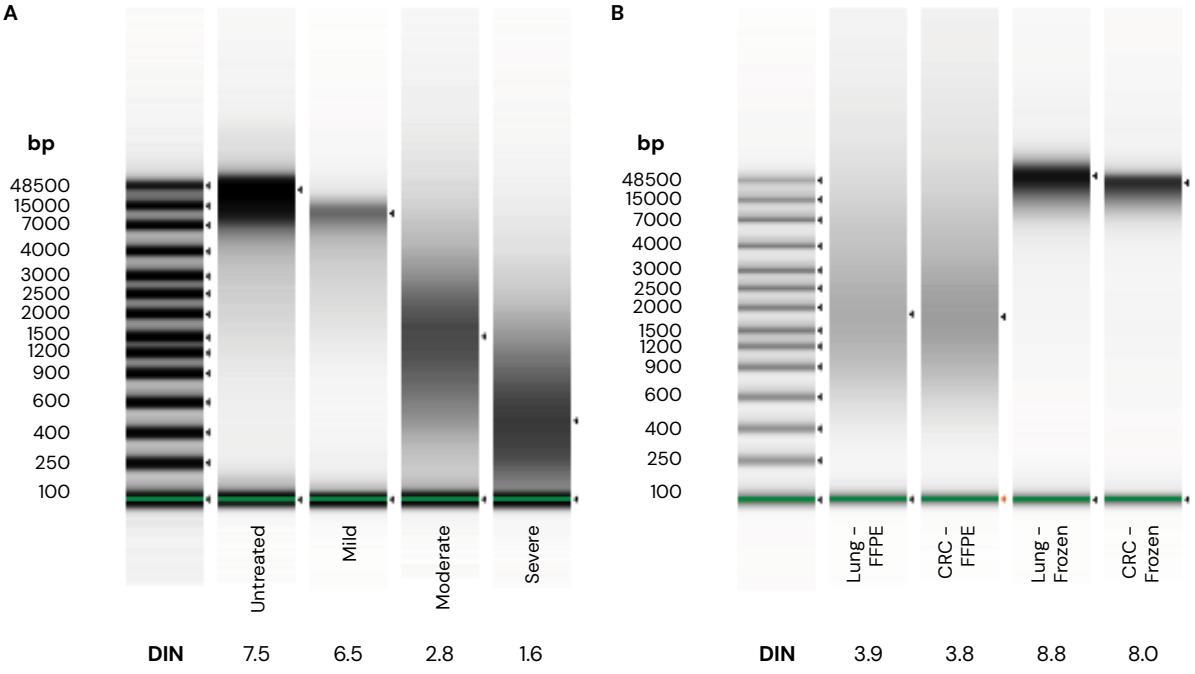


Figure 2: TapeStation gel images for untreated QMRS gDNA, and mild, moderate and severe QMRS fcDNA from Horizon Discovery **(A)**, and matched FFPE and fresh frozen samples for Lung and Colorectal cancers **(B)**.

Triplicates of 10ng and 80ng input fcDNA samples and duplicates of FFPE and fresh frozen DNA at 80ng were prepared using duet evoC. Library quality was assessed using Tapestation and D5000 screentape to determine fragment size and molarity. Figure 3 shows a decrease in yield and fragment size for samples with lower DIN values.

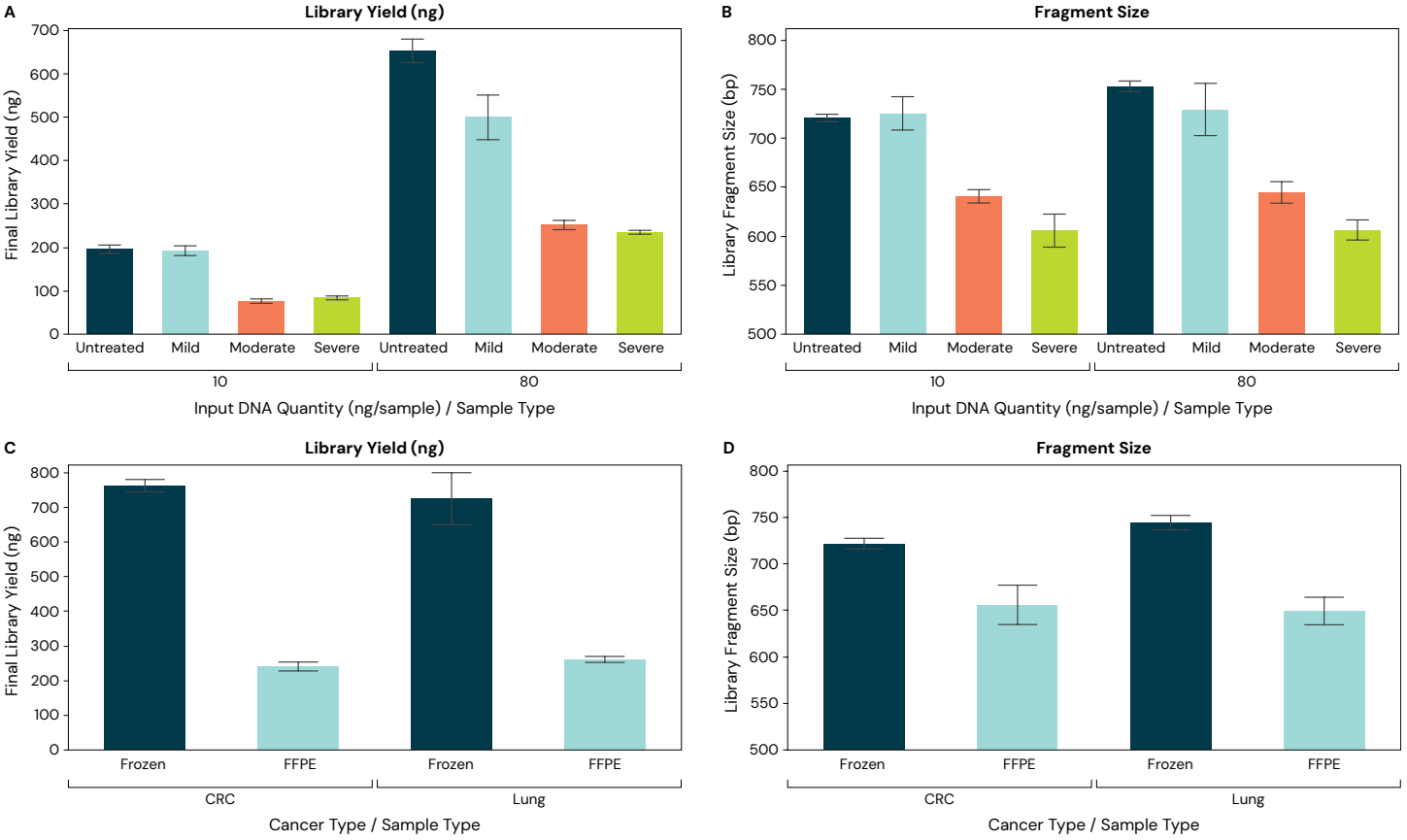


Figure 3: Libraries were prepared using duet evoC and analysed on the TapeStation to determine library yield and fragment size. Library yield **(A)** and Fragment size **(B)** for untreated QMRS and fcDNA libraries. Library yield **(C)** and Fragment size **(D)** for matched FFPE and fresh frozen cancer libraries.

One replicate of each fcDNA sample type (untreated, mild, moderate and severe) and one sample each of Lung and Colorectal matched FFPE and fresh frozen cancer tissues were sequenced using the Illumina NovaSeq 6000 on an S4 flow cell using paired end 2x151 cycles at 30x coverage.

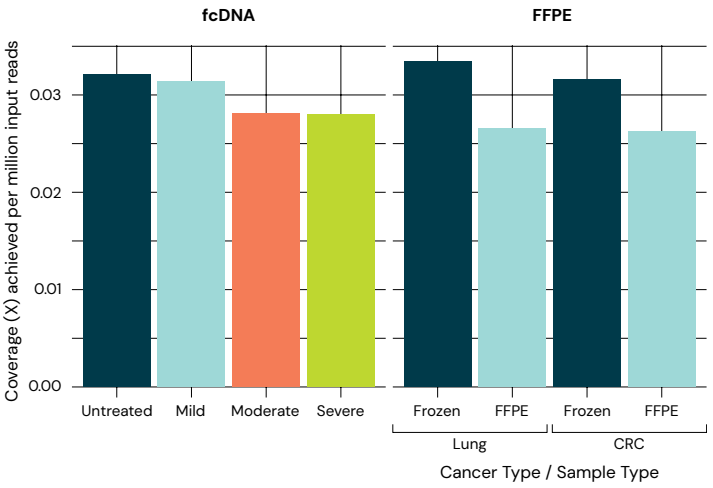


Figure 4: Sequencing coverage (X) achieved per million input reads for all samples.

The coverage yield was calculated for each of the samples sequenced. Figure 4 shows the coverage achieved per million input reads for each sample. Formalin damage resulted in a small drop in coverage, which increased with more severe damage. The FFPE cancer samples also show a small coverage drop compared to the matched frozen equivalents. Although there is a reduction in sequencing yield from FFPE samples, this reduction is small, and the coverage they achieve is still acceptable for high confidence sequencing.

GC coverage plots in Figure 5 show similar distribution between untreated, mild and moderate. Severe damage resulted in reduced representation of low %GC content. GC coverage looks similar between FFPE and equivalent fresh frozen cancer samples.

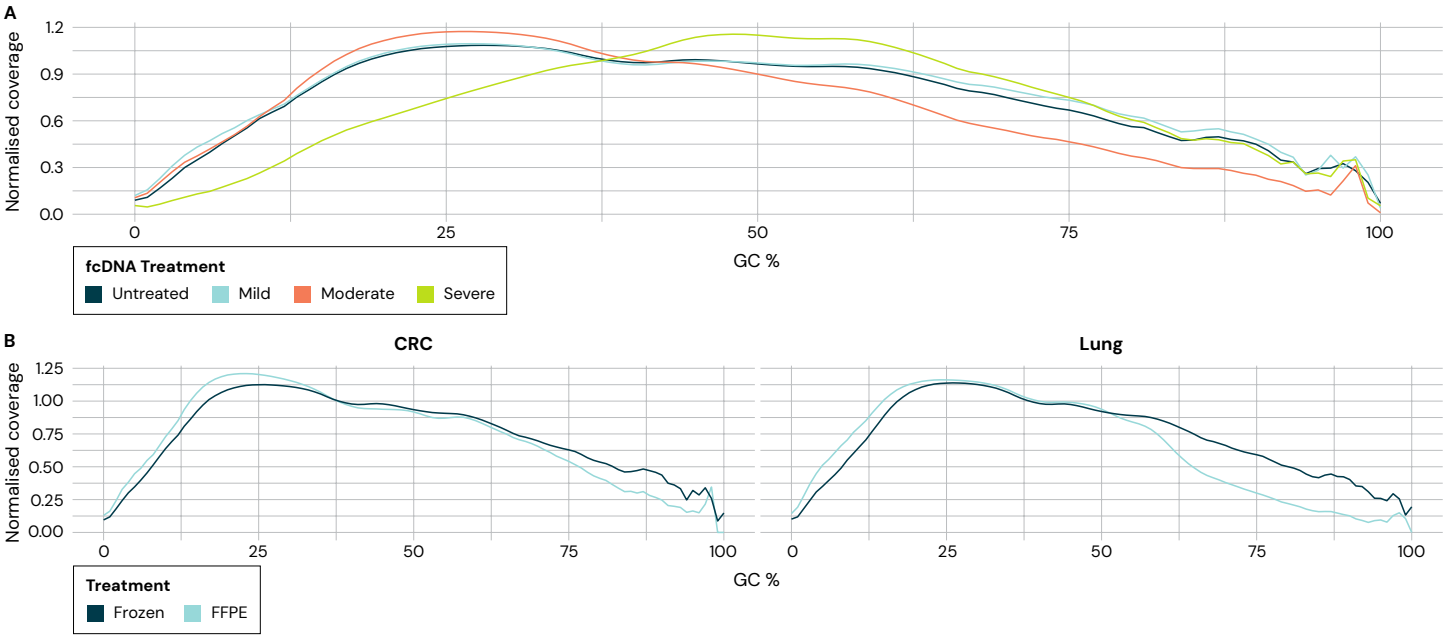


Figure 5: Normalised coverage across different GC% contents in formalin-compromised DNA standards **(A)** and Colorectal and Lung cancer matched FFPE and fresh frozen pairs **(B)**.

To evaluate the impact of formalin damage on epigenetic information, analysis was performed using biomodal's duet analysis suite. Correlation of modC and 5mc levels were measured between untreated controls and fcDNA, and Pearson's r numbers calculated, shown in Figure 6. Correlation is high between all samples, including between untreated and severe fcDNA, suggesting that the impact of the formalin damage on DNA methylation is minimal and that epigenetic information is accurately conserved through the formalin fixation process.

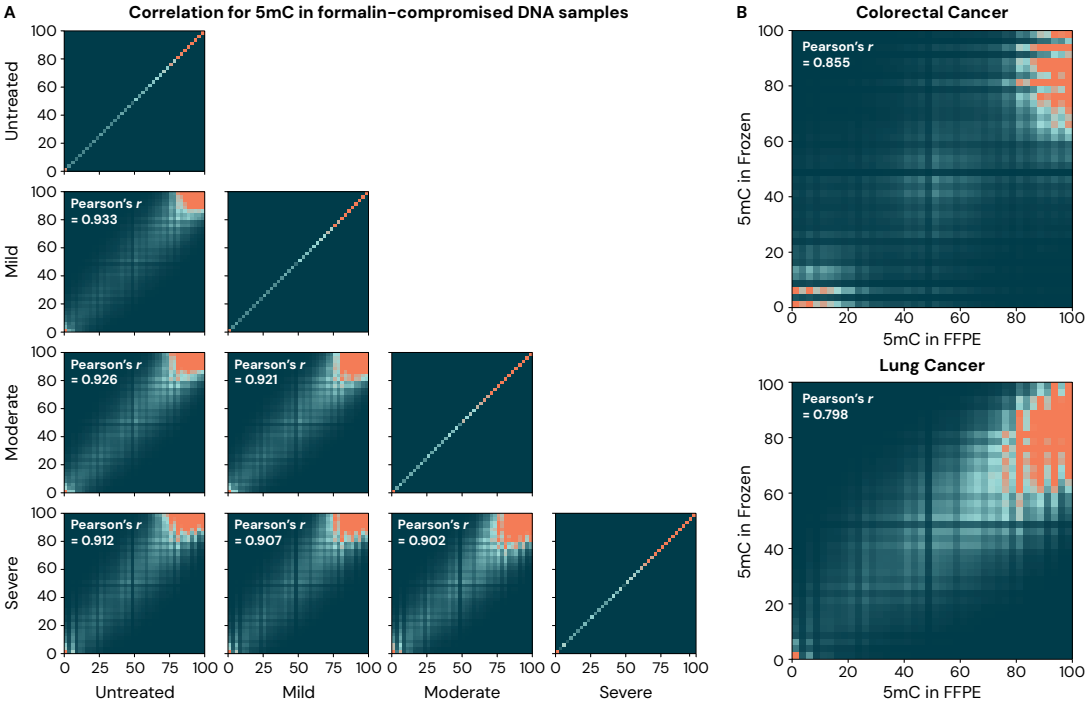
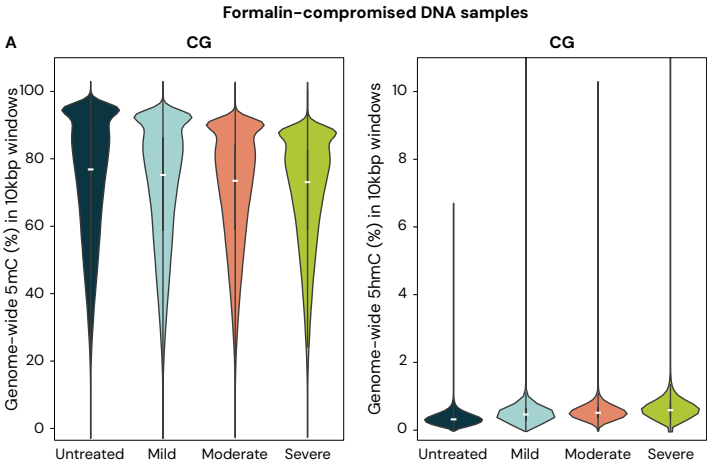


Figure 6: Correlation plots comparing 5mc levels across fcDNA samples **(A)** and matched fresh frozen and FFPE cancer samples **(B)**. All fcDNA samples show high correlation, with Pearson r >0.9 between all samples, including when comparing untreated samples to the most severe formalin damage. Cancer samples also show high correlation between fresh frozen and FFPE conditions.



Genome-wide 5mC levels at CpG contexts decrease with increasing levels of formalin damage, shown in Figure 7 above. Conversely, 5hmC levels at CpG contexts show a small increase with formalin damage. The samples also exhibit a small increase in modC levels at non-CpG contexts associated with formalin damage (not shown).

Figure 7: Percentage of cytosines called as 5mC and 5hmC in CpG contexts.

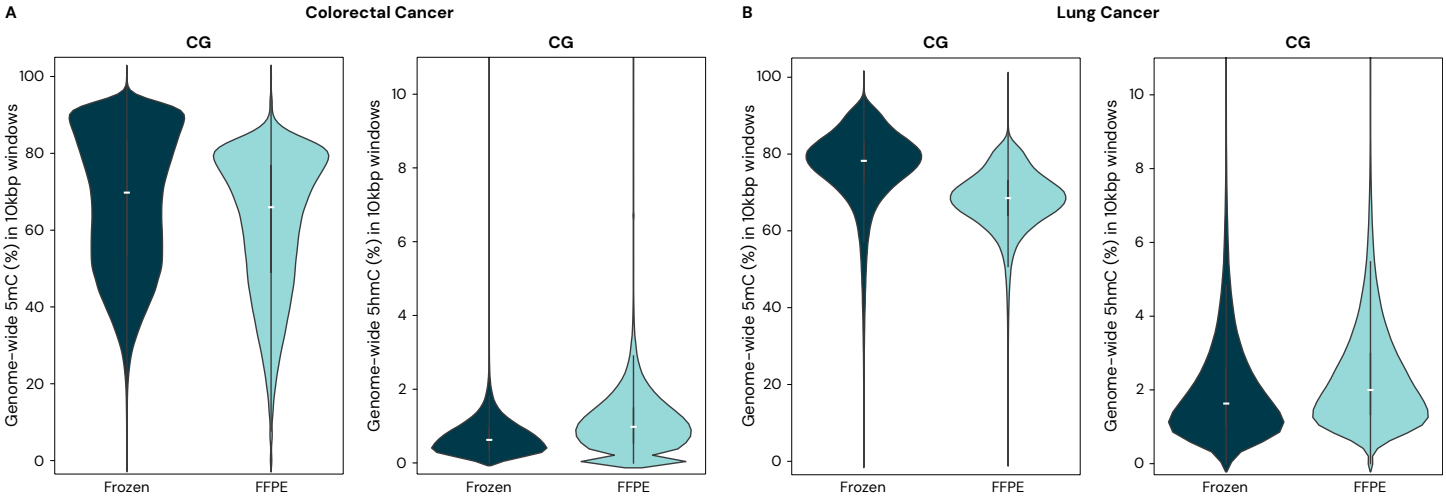


Figure 8: Percentage of cytosines called as 5mC in CpG contexts for Colorectal cancer samples **(A)** and Lung cancer samples **(B)**.

Figure 8 shows the CpG 5mC and 5hmC levels of fresh frozen vs FFPE cancer samples. These exhibit the same trend as the staged formalin damaged samples, showing a genome-wide 5mC decrease and 5hmC increase at CpGs with formalin fixation across both samples.

Conclusion

Here we demonstrate the use of formalin compromised DNA samples as input for the duet multiomics solution evoC workflow. This results in usable libraries that have an expected reduction in fragment size and overall library yield, with the magnitude of the reduction scaling with the relative amount of formalin damage. The %GC coverage trace for these libraries matches closely to untreated samples, with severe damage resulting in a slight reduction in representation of low %GC regions.

Formalin damage is also associated with a small genome-wide reduction in 5mC at CpG contexts, increase of 5hmC at CpG contexts, and increase of 5modC at non-CpG contexts. These small changes again scale with the level of formalin damage. The overall resulting correlation of 5mC calls remains high between all samples from untreated to severe formalin damage. These results suggest that, whilst formalin damage does result in lower library yields and small changes to DNA methylation, the resulting libraries still act as a good representation of the original sample. Changes to DNA methylation are predictable, and the resulting fcDNA 5mC levels still maintain high correlation with an equivalent untreated sample.

We have also demonstrated the compatibility of FFPE cancer samples with duet multiomics solution evoC by comparing matched fresh-frozen and FFPE samples across a range of cancer tissues. These show the same reduction in fragment size and library yield seen with the fcDNA samples, and the %GC coverage trace closely matches between equivalent FFPE and fresh-frozen samples. Small genome-wide 5mC decrease and 5hmC increase is observed in CpG contexts.

FFPE cancer samples are an incredibly important and clinically relevant information source for directly studying the causes and consequences of the disease. Investigating the genetic and epigenetic changes that occur within these tissues is vitally important. From this work, users can be confident that the duet multiomics solution evoC, when used with FFPE samples, gives reliable results that form an accurate representation of the original untreated tissue.

biomodal
Chesterford Research Park
Cambridge, UK
CB10 1XL

+44 (0) 1223 800 700
www.biomodal.com
info@biomodal.com