Exploring transcriptome diversity of human liver using nanopore sequencing

Pedro Garrido-Rodríguez[1], José Padilla-Ruiz[1], María Dolores Frutos-Bernal[2], Rosa Cifuentes-Riquelme[1], María Eugenia de la Morena-Barrio[1], Carlos Bravo-Pérez[1], Vicente Vicente[1], María Luisa Lozano[1], Bruno Ramos-Molina[3], Javier Corral[1], Belén de la Morena-Barrio[1]

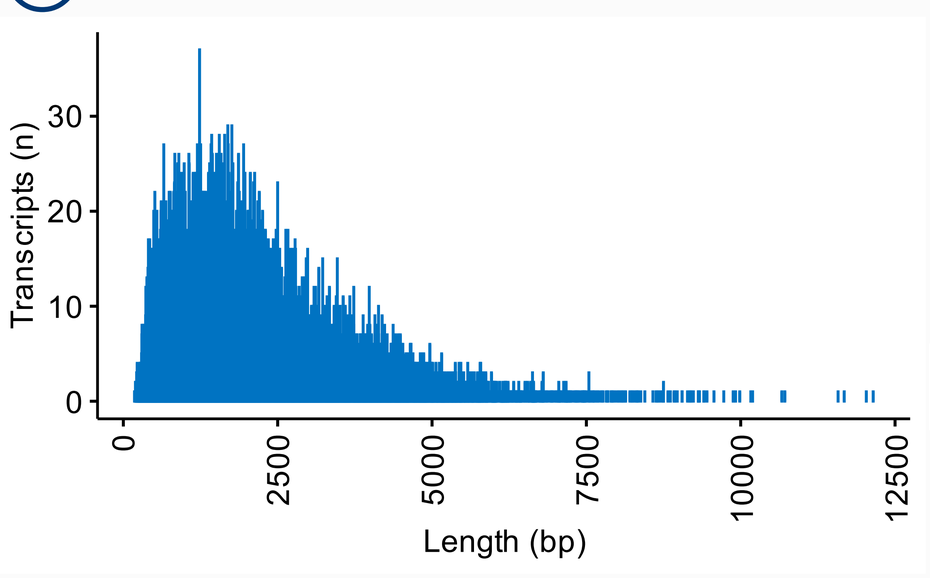
[1]Servicio de Hematología, Hospital Universitario Morales Meseguer, Centro Regional de Hemodonación, Universidad de Murcia, Instituto Murciano de Investigación Biosanitaria (IMIB), CIBERER, ISCIII, Murcia, Spain
[2]Servicio de Cirugía General y Aparato Digestivo, Hospital Universitario Virgen de la Arrixaca, Murcia, Spain
[3]Grupo de Obesidad y Metabolismo, IMIB, Murcia, Spain

INTRODUCTION: Short-read sequencing systems (NGS) are efficient, high-throughput methods routinely used for RNA-Seq. Nonetheless, these show certain drawbacks that could restrict the discovery of novel transcripts. Long-read sequencing (LRS) could improve **transcript identification**, providing further information regarding transcriptome diversity.

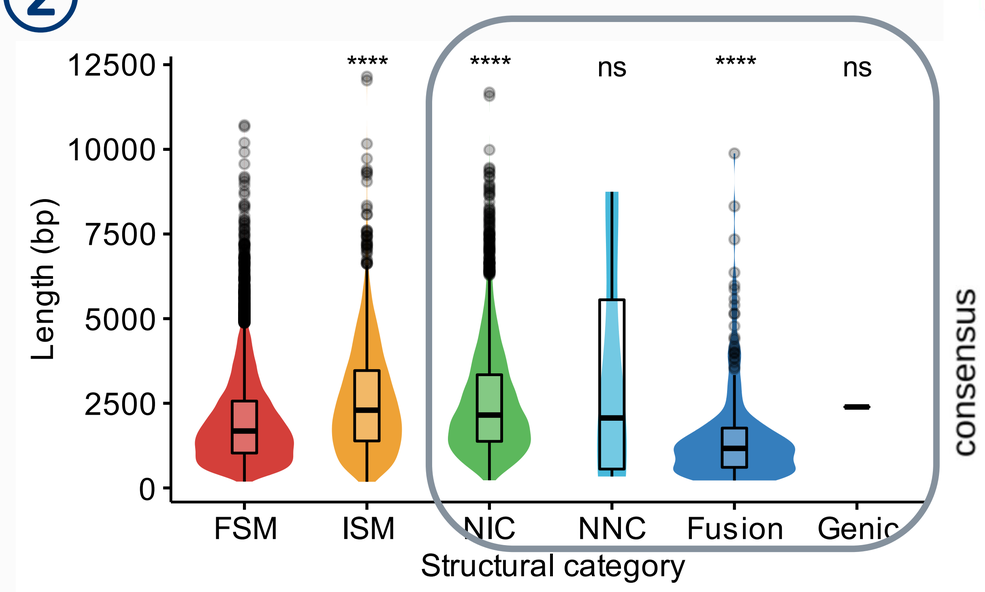
AIM: We explored the potential of nanopore LRS in RNA-Seq studies to define transcriptomic diversity of **human liver**, focusing on the consequences of this diversity in the **hemostatic system**, greatly dependent on hepatic synthesis.

RESULTS: We sequenced 4,8M reads with 789 pb on average ①, assigned to 11155 genes. Remarkably, **4570 new candidate transcripts**, absent on GENCODE, that had similar length than previously reported transcripts, were identified ②. For the hemostatic genes, we found 183 isoforms for 63 genes. 55% of hemostatic genes showed multiple transcripts ③a. Particularly relevant was **fibrinogen gamma (*FGG*)**, with 15 different isoforms ③b **Half of hemostatic transcripts** were **new**. 91 of them were based on known junctions (NIC) ③c Of those, 48 showed intron retention and 43 new combinations of exons. In-silico predictions show only 20 of those new forms would be degraded via NMD. Most transcripts showed low expression levels (114, TPM<10), whilst 69 (22 new) showed medium-high expression levels.

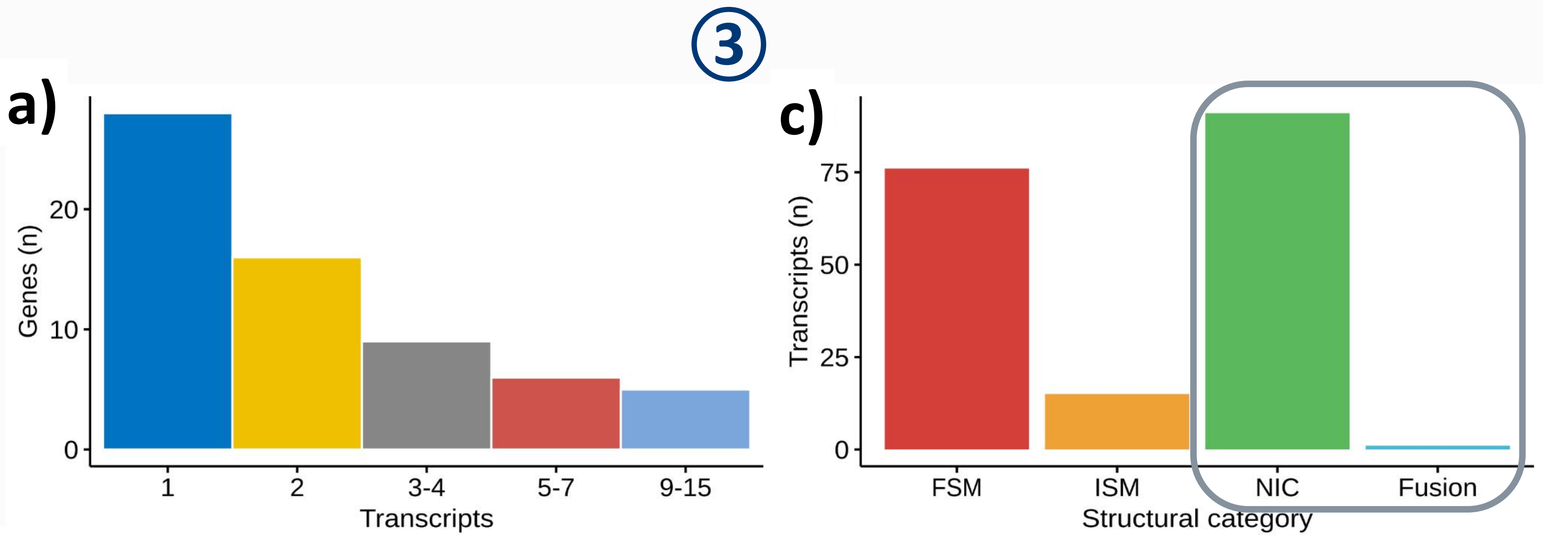
①



②



③



CONCLUSIONS: Nanopore mRNA sequencing **doubles transcriptional diversity** in the hemostatic system compared to conventional NGS methods as it allows identification of new alternative transcripts due to the **direct RNA sequencing** of long reads. These results strongly support an increased proteomic diversity whose physiological and pathological consequences in key biological systems must be explored.

FUNDING: PI21/00174; PMP21/00052; PI20/00505 (ISCIII, FEDER, NG); CP19/00098