TECH NOTE

An optimized solution for cell-free DNA library preparation

- Highly diverse libraries with low duplicate rates
- Unbiased GC coverage
- High-efficiency identification of novel allele variants

Introduction

Cell-free DNA (cfDNA), circulating in blood and found in the plasma component, was first discovered in the 1940s (Mandel and Metais 1948) and has been the subject of renewed attention in the research community due to the easy access to its genetic information. The main source of cfDNA is the apoptotic turnover of hematopoietic cells. DNA fragments are generated by the apoptotic endonuclease caspase-activated DNase (CAD) digesting the chromosomal DNA at regular distances in their nucleosomal arrangement around histones, leading to fragments of various sizes. The cfDNA of primary interest exists as fragments of about 170 bp in length.

This genetic information is being used by translational scientists to better understand the progression of cancer. Circulating tumor DNA (ctDNA) derived from malignant tumors is a component of cfDNA, and libraries prepared from these samples contain genetic information of the tumor (Shaw and Stebbing 2014; Patel and Tsui 2015). For example, Murtaza and colleagues at CRUK performed a research study in which cfDNA libraries were prepared following therapy, and the genetic evolution of several metastatic cancers was followed. One of the major limitations of utilizing next-generation sequencing (NGS) with cfDNA is the difficulty of making sensitive libraries from the relatively low abundance of cfDNA obtained from plasma. Concentrations of cfDNA are quite variable, ranging from 1–20 ng/ml of plasma, and the component of interest is fractionally represented.

SMARTer ThruPLEX technology, which has a history of use in low-input library preparation (Murtaza et al. 2013; Kitzman et al. 2012), has been reformulated and optimized specifically for cfDNA to maximize the library complexity and to preserve the GC representation of the input DNA, with input levels starting at less than 1 ng and ranging to over 30 ng. This new member of the SMARTer ThruPLEX product line, the SMARTer ThruPLEX Plasma-seq kit, is capable of converting cfDNA into high-complexity libraries for Illumina NGS platforms. The three-step, single-tube workflow yields indexed libraries from purified cfDNA within two hours (Figure 1). The generated libraries can be used directly for whole genome sequencing applications or enriched using a custom panel for the leading target enrichment platforms, including Agilent SureSelect and Roche NimbleGen SeqCap EZ.

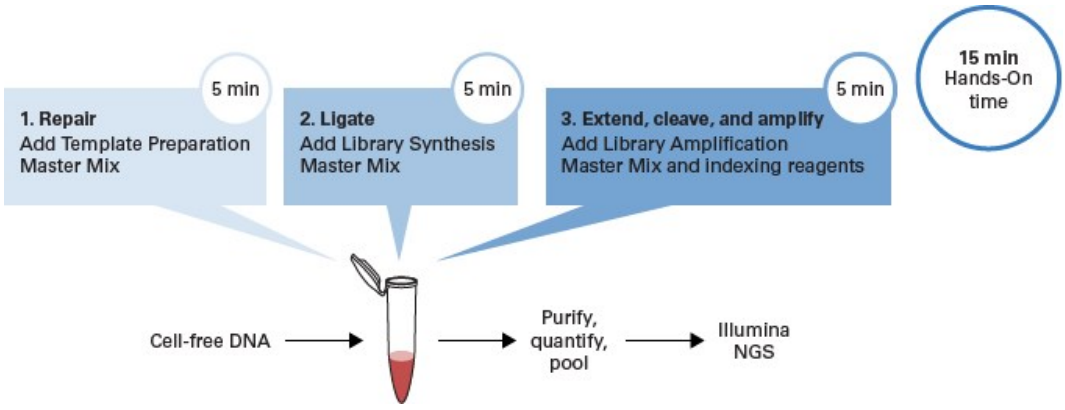

Figure 1. SMARTer ThruPLEX Plasma-seq single-tube workflow.

In the present study, we demonstrate the performance and reproducibility of the SMARTer ThruPLEX Plasma-seq kit in comparison to KAPA Hyper Prep Kit and NEBNext Ultra DNA Library Prep Kit. Furthermore, we show enrichment data that provides a richer view of the genetic variation within the sample.

Results

Preparation of cell-free DNA libraries

There are several library preparation kits for Illumina NGS platforms available, but none have been designed specifically for cfDNA. The SMARTer ThruPLEX Plasma-seq kit can create highly reproducible libraries over a wide input range of cfDNA, from ≤1 ng to 30 ng. Preparation of cfDNA for NGS has usually been done by home-brew kits or kits initially designed to work with mechanically sheared gDNA 200–600 bp in size. Many kits, including the Illumina TruSeq® Nano kit, require a minimum starting amount of 100 ng of DNA, while kits that employ enzymatic fragmentation such as Nextera® DNA Library Prep Kit or KAPA Hyper Plus are not compatible with this type of sample due to the small initial size of cfDNA. In fact, shearing of the cfDNA is unnecessary.

The two kits that were selected for the current test can create libraries from as little as 1 ng (KAPA Hyper Prep Kit) or 5 ng (NEBNext Ultra DNA Library Prep Kit). The SMARTer ThruPLEX Plasma-seq kit is the only kit designed and optimized to efficiently and reproducibly repair, ligate, and amplify NGS libraries from cfDNA. Key to this efficiency and reproducibility for working with DNA fragmented as a result of apoptosis is the use of stem-loop adapters to make libraries, thus eliminating cleanup steps and background problems caused by y-adapters. The SMARTer ThruPLEX Plasma-seq kit also offers several advantages in the workflow when compared to the alternative kits (Table I). Starting with the isolated cfDNA, the SMARTer ThruPLEX workflow creates indexed libraries in a single tube in three steps in about two hours. No sample transfers or intermediate cleanups are necessary. All components including adapters and indexing reagents are provided with the kit, and no optimization is required. Both KAPA Hyper and the NEBNext Ultra have intermediate cleanup steps; both require the purchase of adapters and/or indexing oligonucleotides that often require optimization of concentration to control the number of adapter dimers and other artifacts. Additionally, for low-input amounts of DNA (<25 ng), KAPA recommends optimizing the adapter concentration.

	SMARTer ThruPLEX Plasma-seq	NEBNext Ultra	KAPA Hyper
Recommended input range	1–30 ng	5–100 ng	1–1,000 ng
Total steps	3	4	4
Workflow	1. End repair	1. End repair	1. End repair
	2. Adapter ligation	2. Adapter ligation	2. Adapter ligation
		3. Cleanup	3. Cleanup
	3. Library amplification	4. Library amplification	4. Library amplification
Total hands-on time	15 min	50 min	50 min
Total kit time	~2 hr	~3 hr	~2.7 hr
Sample transfer steps	0	1	1

Table I. SMARTer ThruPLEX Plasma-seq workflow and advantages. The SMARTer ThruPLEX Plasma-seq kit, which includes optimized adapters and indexing reagents, converts cfDNA from plasma samples to indexed NGS libraries in three simple steps in a single tube or well in about two hours; no sample transfer or cleanup steps are required.

Highest diversity and fewest unmapped reads from cfDNA

Libraries created with each of these products were compared on a number of metrics, including library diversity, duplicate reads, and unmapped reads. The SMARTer ThruPLEX Plasma-seq kit yielded significantly higher library diversity while, conversely, a very low percentage of duplicate reads was detected in a low-pass sequencing analysis (Figure 2). The SMARTer ThruPLEX duplication rate was significantly lower than that of the other kits, indicating that with deeper sequencing runs, the SMARTer ThruPLEX Plasma-seq kit would provide more usable data. SMARTer ThruPLEX also had the fewest unmapped reads. These metrics all indicate that the SMARTer ThruPLEX Plasma-seq kit would provide more usable data.

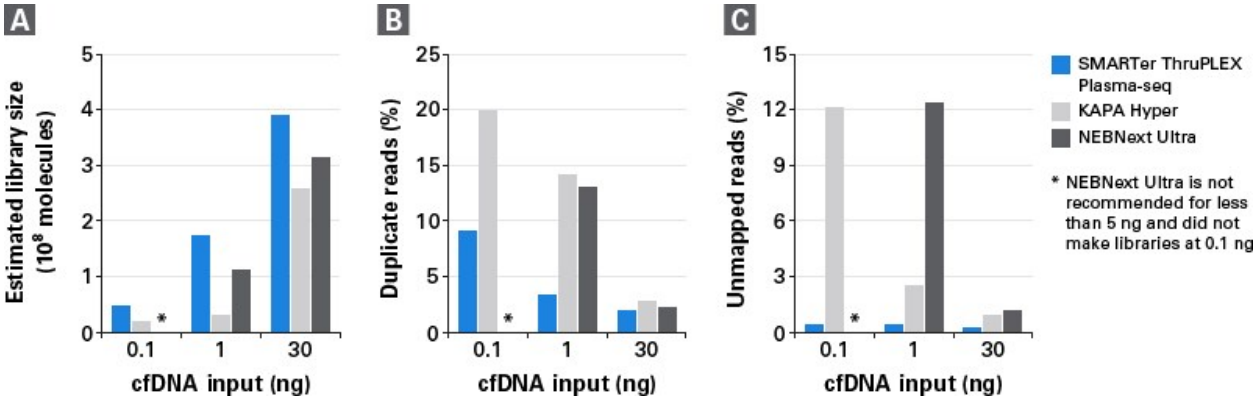

Figure 2. Diverse, reproducible NGS libraries. Libraries created with the SMARTer ThruPLEX Plasma-seq kit yielded more unique molecules (Panel A), fewer duplicate reads (Panel B), and negligible unmapped reads (Panel C). Libraries were sequenced on an Illumina NextSeq® 500 as a paired-end run with 17M to 25M reads per library. Duplication rates were calculated after down-sampling the data to 17M reads per library. Representative data from each sample is shown.

Reproducible, unbiased GC coverage

In GC-bias analysis (Figure 3), the SMARTer ThruPLEX Plasma-seq kit showed well-balanced coverage of the genome between 20% and 70% of GC content. Furthermore, the SMARTer ThruPLEX libraries showed minimal variability across nine individual plasma samples tested. Identical samples were used to prepare libraries with KAPA Hyper, and there was a lack of coverage in the AT-rich region. A separate set of four samples was used to generate libraries for NEBNext Ultra and those, too, lacked the AT coverage. Since the human genome has an average GC content of approximately 42%, libraries prepared with the SMARTer ThruPLEX Plasma-seq kit best represent the original genetic content of the sample.

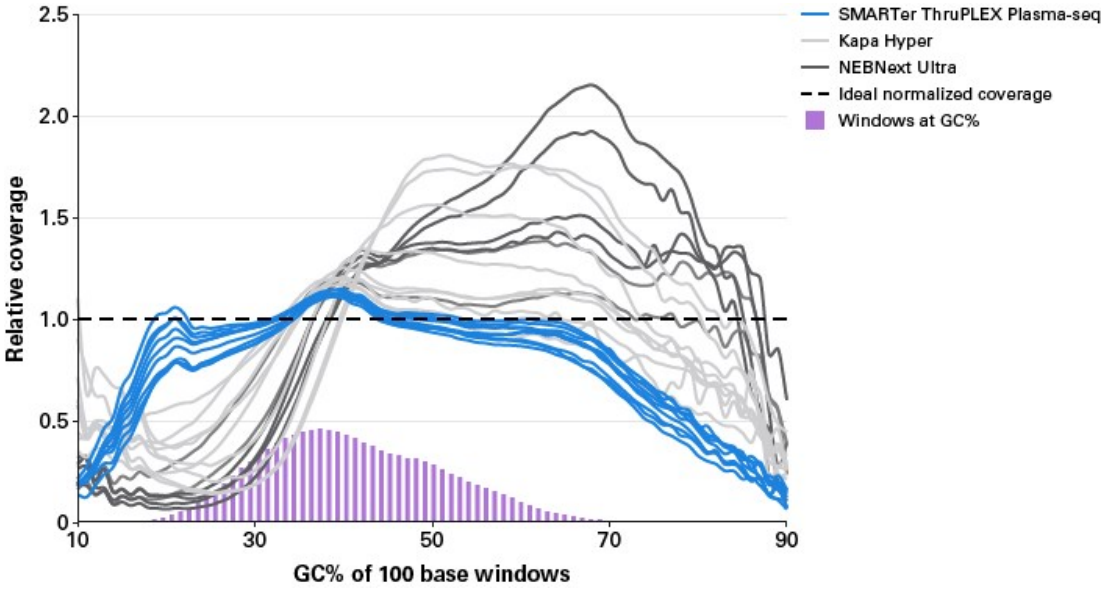

Figure 3. Reproducible, unbiased GC coverage. The SMARTer ThruPLEX Plasma-seq kit provided the most reproducible and unbiased GC coverage across the human genome, showing minimal variability across the nine plasma samples tested. Libraries were prepared from cfDNA isolated from an equivalent of 1 ml of plasma sample and sequenced on an Illumina NextSeq 500. Four separate plasma samples were used to construct the NEBNext Ultra libraries.

Enrichment performance

To better evaluate the performance of the SMARTer ThruPLEX Plasma-seq kit, libraries were enriched using the Agilent SureSelect^{XT2} ClearSeq Human DNA Kinome probe set (Cat. # 5190-4676) according to the SMARTer ThruPLEX SureSelect^{XT2} protocol in the presence of the Universal xGen Blocking Oligos (IDT). Based on approximately 5M total reads for each sample (Table II), a 600-fold enrichment of the human kinome (panel size 3.2 Mbp) was obtained. At 30X coverage, an average of 77% of bases were covered for the cfDNA samples used in this experiment (Figure 4). Using this

data, a highly concordant rate between the replicates for any given sample was found, supporting the ability of the SMARTer ThruPLEX Plasma-seq kit to create libraries that can be used to identify novel allele variants with high efficiency. The identity of the variant calls was confirmed by identifying 98–99% of single-nucleotide polymorphisms (SNPs) in the dbSNP database (Table II). The other 1–2% were novel calls that were generally common to all three replicates of each plasma DNA sample, supporting their biological validity.

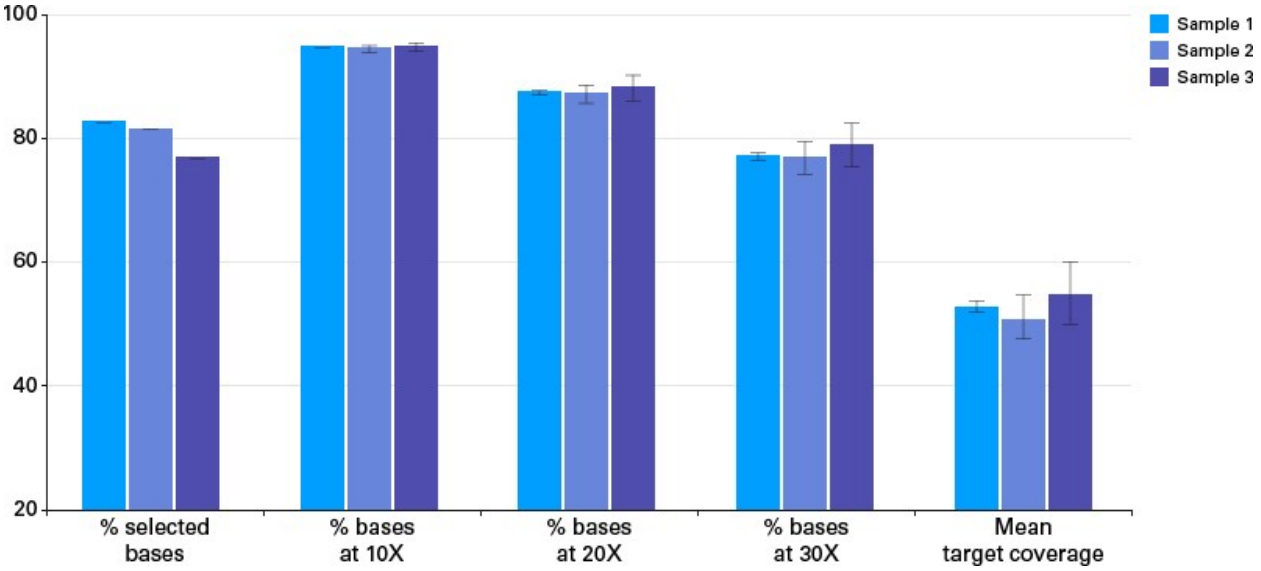

Figure 4. Outstanding target enrichment performance. ThruPLEX Plasma-seq libraries were captured at high efficiency and generated data with deep coverage of the kinome for mutation detection. Libraries were prepared from 3 plasma samples at input amounts of 5 ng, 6.5 ng, and 10 ng in triplicate, and targeted sequencing was carried out on an Illumina MiSeq® using samples enriched with the ClearSeq Human DNA Kinome Panel for SureSelect^{XT2}. On average, 5 M reads were generated per library. Selected bases were successfully captured bases that were in or within 250 bp of the baits.

	Sample A		
	Replicate 1	Replicate 2	Replicate 3
Total reads	4,729,478	4,991,598	4,859,650
Total high-quality uniquely mapped reads	3,309,675	3,645,999	3,392,824
Fold enrichment	645	606	642
Total number of variants identified	1,750	1,792	1,793
Percent of SNPs in dbSNP database	98.9%	98.8%	98.8%

Table II. One SMARTer ThruPLEX Plasma-seq library that was used in kinome capture (Figure 4) was further analyzed for SNP coverage. Results above indicate the number of variants captured and percent of SNPs identified in the dbSNP database are sufficient to allow mutation detection. Libraries were prepared in triplicate from plasma Sample A, enriched using the SureSelect^{XT2} ClearSeq Human DNA Kinome Panel, and sequenced on an Illumina MiSeq platform.

Conclusions

The SMARTer ThruPLEX Plasma-seq kit was specifically developed to produce high-quality libraries from cfDNA. Both the repair and ligation reactions have been reformulated to provide superior results with cfDNA. The optimized repair reaction ensures that the ends of each fragment are blunt and polished to provide high ligation efficiency. Likewise, the ligation reaction has been enhanced for cfDNA molecules to provide maximum ligation of the stem-loop adaptor. The elimination of an intermediate cleanup step and the lack of transfer steps minimize loss of molecules, augmenting the formulation changes to provide this cfDNA-specific product. Our data indicate that the SMARTer ThruPLEX Plasma-seq kit yields better libraries in comparison to its competitors, in terms of diversity, GC bias, and duplicate rates. These libraries are suitable for targeted enrichment and will provide a sensitive tool to allow scientists to easily access and analyze the genetic content of samples from a variety of experimental conditions.

Methods

Plasma sample preparation

Plasma collection was performed by Medical Research Networx, LLC. Blood was collected into BD Vacutainer EDTA tubes and inverted 10 times to mix. Vacutainer tubes were centrifuged (4°C; 12 min; 1,500*g*) with the centrifuge brake off. The plasma layer was then removed, taking care not to disturb the buffy coat, and placed into a 15 ml conical tube. The samples were then centrifuged again (4°C; 12 min; 1,500*g*) before transferring the plasma to a new tube, leaving approximately 0.5 ml to minimize leukocyte carry over. Processed plasma samples were stored at –80°C until DNA was extracted.

Cell-free DNA isolation

Qiagen QIAamp Circulating Nucleic Acid Kit was used according to the manufacturer's protocol without the use of carrier RNA to isolate cfDNA from 5 ml aliquots of plasma samples.

DNA quality control and quantification

Extracted cfDNA eluates from the same individual (15 ml of plasma) were pooled, and the quality of these samples was evaluated on an Agilent BioAnalyzer. The concentration of these samples was measured using Qubit (Thermo Fisher Scientific).

Library preparation

Libraries were prepared from the cfDNA samples following the manufacturer's instructions using the SMARTer ThruPLEX Plasma-seq kit with dual indexes, the NEBNext Ultra DNA Library Prep Kit (New England Biolabs) with dual indexes, and the KAPA Hyper Prep Kit (KAPA Biosystems) with Roche Nimblegen SeqCap EZ adapters diluted to concentrations as recommended in the KAPA protocol for different input amounts. Amplified libraries were pooled and then purified using AMPure XP beads (Beckman Coulter) and eluted in 30 µl of low TE buffer for whole genome sequencing (WGS) or 50 µl of ultrapure water for enrichment. Purified libraries were assessed on the Agilent BioAnalyzer and quantified by qPCR using the KAPA Library Quantification Kit from Bio-Rad Laboratories (KAPA Biosystems). Two WGS experiments and a kinome enrichment were performed (see Table III). For the first, libraries were prepared from three individual plasma samples at input amounts of 0.1 ng, 1 ng, and 30 ng. The amount of mononucleosomal DNA in each sample, as measured by the Bioanalyzer, was 0.09 ng, 0.62 ng, and 15.44 ng. In the second WGS experiment, nine individual plasma samples were tested. cfDNA from a 1 ml aliquot of plasma sample was used to prepare each library; input amounts ranged from 5 ng to 40 ng. For the kinome sequencing experiment, two individual plasma samples were used at input amounts of 6.5 ng and 10 ng, in triplicate.

	Whole genome sequencing				Kinome sequencing	
Samples	Sample 1	Sample 2	Sample 3	Samples 4–12	Sample A	Sample B
Input	0.1 ng	1 ng	30 ng	cfDNA from 1 ml of plasma (5–40 ng)	6.5 ng	10 ng

Table III. Plasma samples and input DNA amount. In the first whole genome sequencing (WGS) experiment, three individual plasma samples were used to construct SMARTer ThruPLEX Plasma-seq libraries at the indicated input amounts. A second WGS experiment used nine individual plasma samples in triplicate. Two separate plasma samples were used for kinome sequencing.

Enrichment

Hybridization and capture of the indexed libraries were carried out using the SureSelect^{XT2} ClearSeq Human DNA Kinome Panel. Briefly, six indexed SMARTer ThruPLEX Plasma-seq libraries, hybridization buffer mix, blocking mix, RNase block, and the ClearSeq Kinome Panel were combined according to the SureSelect^{XT2} protocol. In addition, 1 µl (1 nmol) each of i5 and i7 xGen Universal Blocking Oligo - TS HT (Integrated DNA Technologies) were added into the hybridization reaction which was carried out for 48 hours. Target capture, washes, and final amplification of the enriched libraries were performed according to the SureSelect^{XT2} protocol to obtain captured libraries ready for Illumina sequencing.

Illumina sequencing

Pooled libraries were quantified using the KAPA Library Quantification Kit and loaded onto an Illumina MiSeq or NextSeq 500 flow cell for sequencing. Approximately 17M to 25M reads per library were collected for whole-genome sequencing and 5M reads per library for kinome sequencing.

Data analysis

Sequences were analyzed on the DNANexus platform. Reads were aligned to the human genome, hg19, using the Burrows-Wheeler Algorithm, BWAMEM6, to generate BAM files. For WGS data, reads were first down-sampled to equal numbers across all samples. Down-sampled BAM files were assessed using Picard Mark Duplicates7 to count duplicate reads and estimate diversity (estimated library size), and Picard Collect GC Metrics was used to determine biases based on sequence GC content. For kinome sequencing data, after mapping with BWA-MEM, Picard CalculateHsMetrics was used to determine capture quality metrics. For SNV analysis, Agilent SureCall was used to identify variants within the targeted exons of the kinome, and Illumina Variant Caller was used to annotate variants.

References

Broad Institute. Picard Tools - A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. <http://broadinstitute.github.io/picard/>

Kitzman, J. O. *et al.* Noninvasive Whole-Genome Sequencing of a Human Fetus. *Sci. Transl. Med.* 4, 137ra76–137ra76 (2012).

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009).

Mandel, P. & Metais, P. Les acides nucléiques du plasma sanguin chez l'homme. *C. R. Seances Soc. Biol. Fil.* 142, 241–3 (1948).

Murtaza, M. *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 497, 108–112 (2013).

Patel, K. M. & Tsui, D. W. Y. The translational potential of circulating tumour DNA in oncology. *Clin. Biochem.* 48, 957–961 (2015).

Shaw, J. A. & Stebbing, J. Circulating free DNA in the management of breast cancer. *Ann. Transl. Med.* 2, 3 (2014).

Related Products

Cat. #	Product	Size	License	Quantity	Details
R400491	SMARTer® ThruPLEX® Plasma-seq 48S Kit	48 Rxns			
R400492	SMARTer® ThruPLEX® Plasma-seq 96D Kit	96 Rxns			
R400490	ThruPLEX® Plasma-seq 12S Kit	12 Rxns			

Add to Cart

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999