1 **Genetic analysis of sudden unexpected death cases: Evaluation of library preparation methods**

2 **to handle heterogeneous sample material**

3

4 **Stefanie Scheiper-Welling[1,2], Stephanie Körber[2], Christof Geisen[2], Marcel A. Verhoff[1], Silke**

5 **Kauferstein[1]**

6 [1] Institute of Legal Medicine, Goethe University, University Hospital Frankfurt, Frankfurt, Germany

7 [2] German Red Cross Blood Center, Institute of Transfusion Medicine and Immunohaematology,

8 University Hospital Frankfurt, Frankfurt, Germany

9

10 To whom correspondence should be addressed: Stefanie Scheiper-Welling, Institute of Legal Medicine,

11 University of Frankfurt, Kennedyallee 104, D-60596 Frankfurt am Main

12 E-mail: scheiper@med.uni-frankfurt.de

13

Abstract

14 Over the past years, next-generation sequencing (NGS) technologies revolutionized the possibilities in

15 a broad range of application areas. Also in the field of forensic genetics, NGS continuously gained in

16 importance and attentiveness. A significant number of sudden cardiac deaths (SCD) in the young is

17 due to heritable arrhythmia syndromes emphasizing the need of examining the genetic basis in these

18 cases also with regard to the identification of relatives and/or patients being at risk. As a result, high-

19 throughput methods became of increasing value in molecular autopsy investigations enabling the

20 analysis of a broad spectrum of genes.

21

22 Most standard protocols are optimized for high-quality samples and frequently not directly applicable to

23 challenging forensic sample material. In the present study, we intended to examine a comprehensive

24 gene panel associated with SCD and inherited arrhythmogenic disorders. We compared three different

25 hybridization-based library preparation technologies in order to implement a suitable NGS workflow for

26 heterogeneous, forensic as well as diagnostic sample material.

27

28 The results obtained indicated, that the Illumina technologies Nextera DNA Flex and TruSeq were

29 compatible with samples exhibiting varying levels of degradation. In comparison, the TruSight method

30 also resulted in good sequencing data, but seemed to be more dependent on DNA integrity. The

31 preparation protocols evaluated in our study are not restricted to molecular autopsy investigations and

32 might be helpful for and transferrable to further forensic research applications.

33

34 **Keywords:** Sudden Cardiac Death (SCD), DNA integrity, Degradation, Next-generation sequencing,

35 Targeted sequencing

36

1. Introduction

Over the past years, next-generation sequencing (NGS) technologies rapidly evolved as an indispensable tool in a wide spectrum of research applications, in clinical diagnostics as well as forensic genetics offering a lot of new possibilities [1, 2]. For instance, NGS became of increasing importance in molecular autopsy investigations [3–5]. A significant number of sudden cardiac death (SCD) cases - especially in the young - was found to have its origin in inherited arrhythmia syndromes, which include primary electrical heart disorders as well as cardiomyopathies [6, 7]. These diseases have in common that they are phenotypically and genetically heterogeneous [8]. As sudden death often reveals the first and only sign of a hereditary disease, elucidating its genetic basis in victims of sudden unexpected death (SUD), in their relatives and even in patients presenting clinical abnormalities, NGS enables a fast, cost-efficient and simultaneous analysis of a high number of genes [1, 3, 5, 9]. Genetic analyses in these cases allow to initiate cascade screening in affected families in order to identify relatives at-risk enabling the implementation of preventive measures.

In the field between forensic genetics and molecular diagnostics it is important to perform genetic screening and subsequent interpretation of the results as close as possible to the diagnostic standards as the analyses might implicate far-reaching consequences for relatives and patients [10]. Therefore, it is important to meet the quality requirements implementing an NGS workflow for molecular autopsy and subsequent family investigations. During the last years, various high-throughput techniques and an increasing number of sample preparation methods have been developed [1, 2].

For targeted sequencing, every method has its advantages and disadvantages. In forensic applications, PCR-based target enrichment is by far the most prevalent and sensitive method [11]. Furthermore, amplicon-based technologies may be more suitable and may result it better sequencing performance when working with limited sample amounts, smaller panels or targets exhibiting high homologies. In contrast, hybridization capture-based enrichment methods are more frequently used in molecular diagnostics and generally have advantages when analyzing e.g. larger gene panels or even whole exomes [2, 11–14].

In the present study, a comprehensive panel of candidate genes associated with inherited arrhythmogenic diseases and sudden cardiac death to be analyzed by next-generation sequencing was examined. As postmortem samples show high variability in quality and DNA integrity, most standard NGS protocols are often not directly applicable, because they are optimized and standardized for the analysis of intact, high-quality genomic DNA (gDNA). Thus, for the implementation of NGS, we compared different workflows, which were adapted to analyze heterogeneous sample material. Three library preparation technologies were tested and evaluated for its practical application close to diagnostic standards.

2. Material and Methods

2.1 Ethical statement

The present study was approved by the Ethical Commission of the University Hospital, Goethe University of Frankfurt (protocol number E84/06).

2.2 Sample preparation

Samples from sudden unexpected death cases and from patients exhibiting clinical abnormalities pointing towards an arrhythmogenic disease were used to test three library preparation technologies for subsequent next-generation sequencing. DNA was extracted from either blood (n=48), renal (n=1), pulmonary (n=12) or muscle (n=1) tissue, cardiac FFPE (formalin-fixed paraffin-embedded, n=1) samples as well as from cardiac or kidney tissue samples stored in formalin (n=3). Blood samples were extracted applying the Maxwell® RSC Blood DNA Kit (Promega, Madison, USA) or the NucleoSpin® Tissue Kit (Macherey Nagel, Düren, Germany), which was also used for tissue samples. DNA from FFPE samples was extracted using the AllPrep DNA/RNA FFPE Kit (Qiagen, Hilden, Germany). Extractions were performed according to the manufacturers' protocols. Samples stored in formalin only were rinsed overnight and were processed using the same kit, starting from step 5 after deparaffinization.

Following DNA extraction, sample purity and quantity were assessed using the Nanodrop® ND-1000 Spectrophotometer v3.1.0 (Intas, Göttingen, Germany) and Qubit 3 Fluorometer in combination with the dsDNA BR and HS assay kits (Invitrogen, Carlsbad, USA), respectively. Only DNA samples showing high purity (A260/A230 ratio generally > 2, at least > 1.8) were used for subsequent sequencing. DNA integrity was examined using genomic DNA or D1000 ScreenTape assays and Agilent 4200 TapeStation (Agilent Technologies, Santa Clara, USA).

2.3 Library preparation and targeted sequencing

Paired-end libraries were prepared using the TruSight cardio panel (Illumina, San Diego, USA) consisting of 174 genes associated with cardiac diseases (supplemental data) and three different library preparation technologies of the company Illumina. Library preparation was performed applying either the TruSight Cardio Kit (bundled solution), the TruSeq Exome Kit - where exome oligos were replaced by TruSight cardio oligos, or the relatively new Nextera Flex technology (now referred to as Illumina DNA prep) as follows:

TruSight technology: Libraries were prepared using the TruSight Cardio Sequencing Kit according to the manufacturer's manual. Since the standard protocol is optimized for intact gDNA samples of high

108 quality, the volume of the 'tagment DNA enzyme' for tagmentation was reduced to 10 µl to avoid over-
109 fragmentation and thus smaller insert sizes.

110

111 TruSeq technology: Library preparation was performed using the TruSeq Exome Kit. As mentioned
112 above, exome oligos were replaced by TruSight cardio oligos (Illumina). The instructions of the kit's
113 protocol were followed. Certain steps were adjusted according to the Illumina TruSeq technical note for
114 the library preparation with regard to FFPE samples to achieve better results with heterogeneous
115 samples. For samples exhibiting a DNA integrity number (DIN) lower than 6, 300 ng of DNA, for samples
116 showing a higher DIN (≥ 6), 200 ng DNA were used (instead of recommended 100 ng DNA). DNA was
117 sheared by means of sonication using Covaris M220 Focused-ultrasonicator (Covaris, Woburn, USA).
118 In the PCR program 'PCR nano' of theTruSeq reference guide, cycles were increased from 8 to 12
119 cycles. For enrichment, 500 ng instead of 100 ng (for 12-plex) were used to pool DNA libraries. Enriched
120 libraries were amplified in 13 instead of 8 cycles.

121

122 Nextera Flex technology: Libraries were prepared following the manufacturer's protocol (Nextera™ Flex
123 for Enrichment). Since the samples showed varied considerably regarding their DNA integrity,
124 recommendations for formalin-fixed paraffin-embedded samples were followed as described in the
125 protocol. Of samples exhibiting a DIN lower than 5, 300ng of DNA, above a DIN of 5, 200ng were used
126 for tagmentation. In order to obtain better performance in GC-rich regions, the PCR program 'Amplify
127 tagmented DNA' was modified and denaturation times in steps 2 and 3 were increased to 4 min., 30 s
128 instead of 3 min., 20 s, respectively). For enrichment, 500 ng per pre-enriched library were used.
129 Hybridization of probes was performed overnight. Enriched libraries were amplified with adjusted
130 doubled denaturation times in steps 1 and 2 of the protocol.

131

132 Concentration and quality of pre-enriched and enriched libraries were checked fluorimetrically and by
133 applying D1000 and High Sensitivity D1000 TapeStation assays (Agilent Technologies), respectively,
134 as mentioned above. A genomic DNA reference sample (e.g. NA12878, Coriell Institute) was included
135 in every library preparation process and served as a control in each sequencing run. Sequencing was
136 carried out on Illumina platforms MiSeq or MiniSeq System (2x150 bp paired end reads, using v2
137 reagent kit or high output kit, respectively). The resulting reads were aligned to the human reference
138 genome GRCh37/hg19. Evaluation of the data was performed using Illumina Analysis Software and
139 Illumina Sequencing Analysis Viewer (Illumina) as well as GensearchNGS software (PhenoSystems,
140 Braine le Chateau, Belgium). Next to this, only NGS runs meeting the following quality criteria were
141 considered as good sequencing runs: Cluster densities should be close to the manufacturer's
142 recommended range, a quality score of Q30 should be reached by at least 90 % of the bases and at
143 least 98 % of core genes should be covered ≥ 20x.

144

3. Results

3.1 Integrity of starting material

The integrity of the input gDNA may critically affect the success of library preparation and subsequent sequencing. Therefore, quality control of the starting material was performed and the gDNA integrity of samples, that showed good purity and quantity following extraction, was assessed. Depending on sample origin, the kind of sample material and its storage before extraction, the extracts exhibited high variability of gDNA integrity ranging between DNA integrity number (DIN) values of 1 to 9.1, as shown in figure 1.
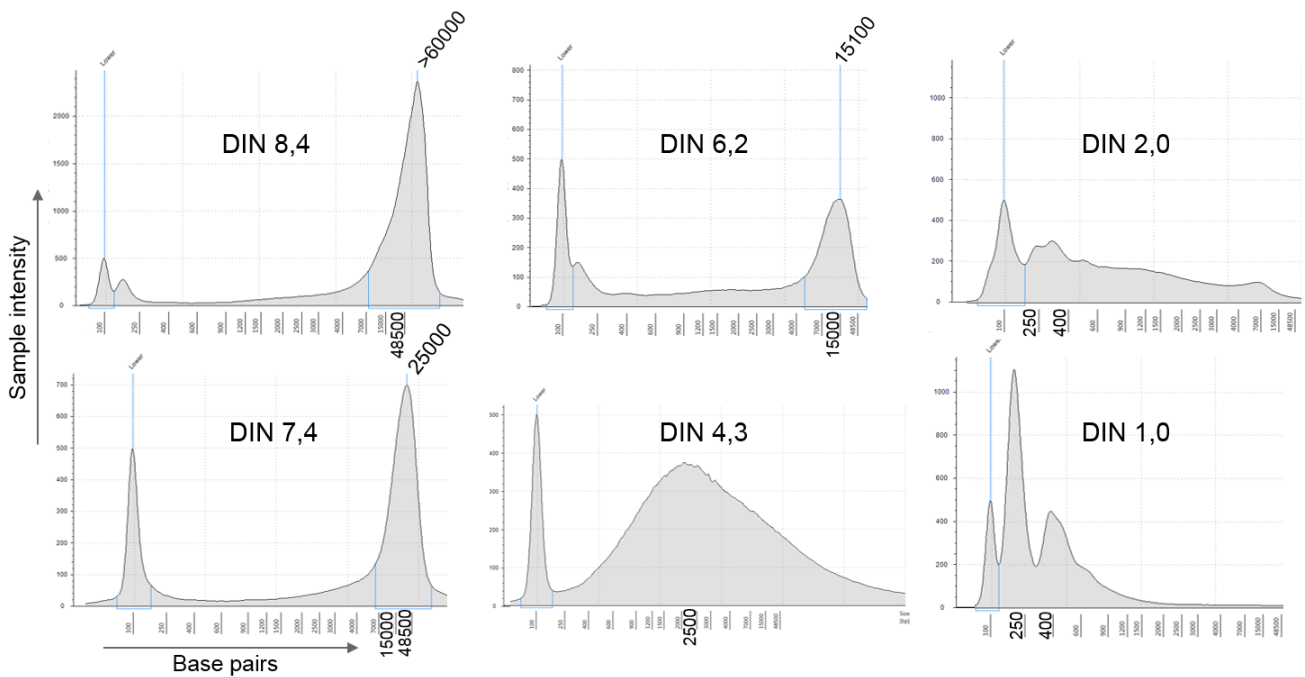
Figure 1: Samples showed high variability of gDNA integrity. Representative electropherogram patterns using the 4200 TapeStation system and Genomic DNA Screen Tape assay display the range of gDNA integrity of the samples (DIN scale 1-10, where a high DIN indicates highly intact gDNA; sample intensities indicated in normalized fluorescent units).

Diagnostic samples revealed DIN values between 6.1 and 9.1. Therefore, extracts with values above DIN 6 were considered to contain relatively intact DNA. As shown in figure 1, samples with a DIN value > 6 exhibit one well-defined peak only, whereas samples with DIN close to 6 already display signs of partial degradation, but there is still a major peak visible, which is shifted towards smaller sizes.
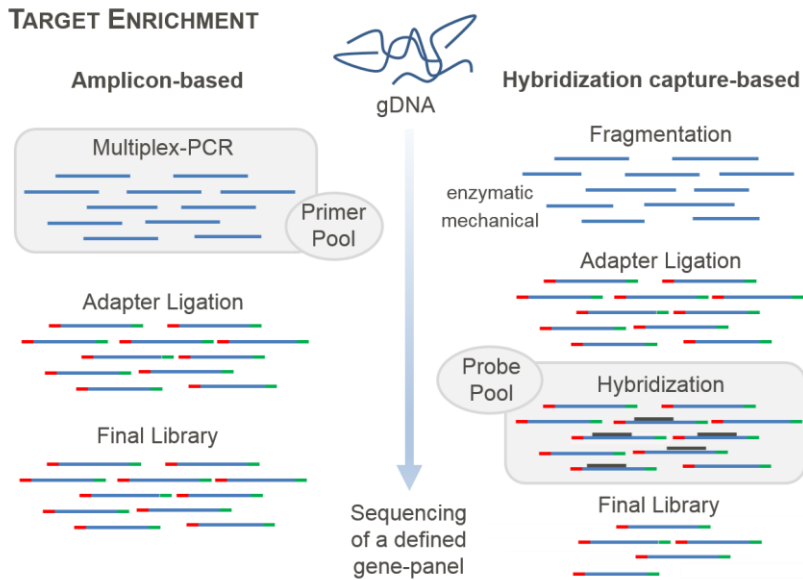
169 **3.2 Approach for targeted sequencing**

170 For the present scope of application, we considered a commercially available Illumina-compatible
171 hybridization capture-based enrichment workflow (figure 2) to be most suitable for examining a broad
172 spectrum of genes associated with inherited arrhythmia diseases.

173

174 **Figure 2: Common enrichment options for targeted sequencing.** Using amplicon-based methods, targets are
175 enriched by PCR amplification in e.g. one or multiple multiplex reactions. With hybridization-capture enrichment
176 methods, gDNA is first fragmented. Afterwards, targets are selectively captured by hybridization of specific
177 biotinylated probes complementary to the regions of interest and subsequently recovered using streptavidin-
178 magnetic beads.

179

180 The targeted gene capture panel 'TruSight Cardio' of Illumina comprised the most relevant genes
181 enabling the analysis in a single sequencing assay. Based on the results of the DNA integrity
182 measurements the following approach for targeted sequencing of heterogeneous sample material was
183 established using three different library preparation technologies (Figure 3) according to the protocols'
184 requirements:
185 We initially started with the Illumina TruSight technology (TruSight Cardio Sequencing Kit, bundled
186 solution). This kit is based on enzymatic fragmentation. During 'tagmentation', the DNA is fragmented
187 and tagged with adapter sequences in a single step. However, enzymatic fragmentation is highly
188 sensitive to the amount and quality of input nucleic acid. Hence, protocols are difficult to adapt to
189 applications with samples exhibiting varying levels of degradation. Therefore, we decided to apply this
190 technology to gDNA samples exhibiting a DIN value > 6.
191 In a second step and for best possible processing of samples showing lower levels of DNA integrity,
192 the Illumina TruSeq technology was tested. This procedure is based on mechanical DNA shearing.
193 Using this method, accurate gDNA fragmentation is less dependent on the concentration and integrity

of the starting material. This technology was mainly applied to highly degraded samples with a DIN < 6.

Subsequently, we applied the new library preparation method Nextera DNA Flex (now named Illumina DNA Prep) employing enzymatic 'on-bead tagmentation'. This technology promised to be compatible with a wide range of input types and amounts. Next to this, it should also be suitable for degraded as e.g. FFPE samples. This kit was used independently of the sample integrity.

In order to obtain appropriate insert sizes and final enriched libraries, all protocols were adapted to the use of heterogeneous material as described in the methods section.
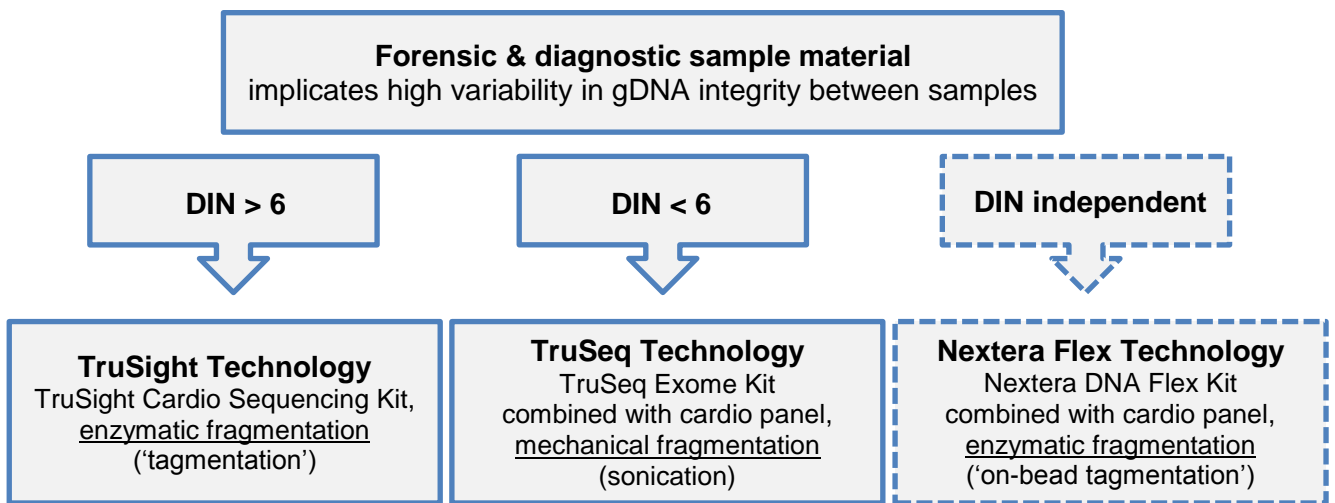

Figure 3: Approach performing genetic analysis of heterogeneous sample material using Next-generation sequencing. Intact samples displaying a relatively high gDNA integrity (DIN > 6) were processed with the TruSight technology. Samples exhibiting high degradation levels (DIN < 6) were prepared using the TruSeq protocol. The new Nextera DNA Flex technology was applied DIN independently to all sample types.

3.3 Comparison of library preparation methods and sequencing performances

For each sequencing run, libraries of 12 samples were processed, pooled and enriched using the TruSight cardio panel for subsequent analyses. Library preparation was performed with either the TruSight, TruSeq or Nextera DNA Flex technology. By adapting the protocols, uniform libraries of proper size and concentration were obtained. Table 1 represents the results of two NGS runs for each preparation technology. Excluding the reference sample, the average DNA integrity number of samples prepared with the TruSight method was 7.8 in the first and 6.9 in the second library preparation workflow. Using the TruSeq technology, the samples exhibited lower values, namely DIN 4.1 in the first and 2.9 in the second preparation workflow, and samples prepared with the Nextera DNA Flex workflow DIN values of 4.7 and 5.9, respectively. Only pre-enriched libraries conforming to the quality requirements were further processed.

In each NGS run, cluster densities within an adequate range were observed. The sequencing data were of good quality, as a quality score of Q30 was obtained by > 91 % of the bases (averaged 93.3 % ± 1.3, n=6). Regarding the sequencing performance for the entire panel consisting of 174 genes with known cardiac associations, only samples showing 20x coverage in at least 98 % of the targets (without duplicates) were considered for evaluation (table 1). All the technologies tested were found to provide good coverage depths, target coverage and uniformity of coverage. However, it has to be considered, that only comparatively intact samples were processed using the TruSight protocol.

Table 1: **Comparison of Next-generation sequencing performances using three different library preparation technologies for genetic analysis of heterogeneous sample material.** Average mean coverage depth, coverage uniformity and target coverage (20x, 50x) including standard deviations are shown. For target coverage statistics, duplicate reads were excluded. For each technology, the results of two NGS runs are listed. The third row of each technology section summarizes the results of both runs. Further evaluation parameters represented the assay time and the equipment as well as the amount of DNA required.

Technology	Mean Coverage Depth	Uniformity of Coverage (pct > 0.2 *mean)	Target Coverage 20x	Target Coverage 50x	Assay Time	Equip-ment	Input (ng)*
TruSight enzymatic fragmentation	314	98.21 % (± 0.31)	99.75 % (± 0.32)	98.08 % (± 2.58)	+/-	+/-	50
	356	95.89 % (± 1.19)	99.56 % (± 0.3)	96.34 % (± 2.71)			
	335	97.05 % (± 1.46)	99.65 % (± 0.32)	97.21 % (± 2.73)			
TruSeq mechanical fragmentation	407	99.06 % (± 0.12)	99.89 % (± 0.03)	99.62 % (± 0.32)	-	-	200 - 300
	925	99.16 % (± 0.16)	99.9 % (± 0.00)	99.83 % (± 0.07)			
	654	99.11 % (± 0.15)	99,9 % (± 0.02)	99.72 % (± 0.25)			
Nextera DNA Flex enzymatic fragmentation -'on-bead tagmentation'	555	99.15 % (± 0.78)	99.88 % (± 0.04)	99.64 % (± 0.15)	+	+/-	flexible range 50-300 tested
	667	98.65 % (± 0.37)	99.84 % (± 0.07)	99.55 % (± 0.28)			
	611	98.9 % (± 0.65)	99.86 % (± 0.06)	99.59 % (± 0.23)			

*Values concerning input amounts refer to initial gDNA inputs successfully tested in this study and may differ from values stated in corresponding protocols.

Sequencing coverage depth and uniformity were higher for libraries prepared using the TruSeq and Nextera DNA Flex workflow compared with data obtained using the TruSight technology. The TruSeq technology showed the lowest deviations in uniformity and target coverage and the best target coverage (20x as well as 50x) over the entire gene panel, followed by the Nextera DNA Flex and the TruSight workflows. Slight variations in coverage may have noticeable impact on the evaluation of the sequencing data, especially when analyzing the coverage of core genes.

3.4 Coverage of Core Genes

The entire and sufficient coverage of core genes is one of the most important aspects in molecular genetic analyses. Particularly in molecular autopsy investigations, this are the main prevalent genes associated with arrhythmogenic disorders as Long-QT Syndrome (LQTS), Brugada Syndrome (BrS), Short-QT Syndrome (SQTS), Catecholaminergic Polymorphic Ventricular Tachycardia (CPVT) as well as hypertrophic, dilative and arrhythmogenic cardiomyopathy (HCM, DCM, ACM, respectively) according to the current guidelines [15]. Therefore, the coverage of 15 major candidate genes associated with SCD was analyzed (table 2). The evaluation was based on the coverage of coding regions including ten base pairs of the flanking intronic regions.

Table 2: Coverage of 15 core genes associated with sudden cardiac death (SCD) and arrhythmogenic disorders using three different library preparation technologies. Coverage analyses was based on the genes *SCN5A*, *KCNQ1*, *KCNH2*, *KCNJ2*, *RYR2*, *MYH7*, *MYBPC3*, *TNNI3*, *TNNT2*, *LMNA*, *BAG3*, *PKP2*, *DSC2*, *DSG2* and *DSP*. Average values for each run and technology, respectively, are shown.

Technology	Coverage of core genes 20x	Coverage of core genes 50x
TruSight enzymatic fragmentation	99.88 % (± 0.26 %)	98.58 % (± 2.17 %)
	99.82 % (± 0.26 %)	97.55 % (± 2.05 %)
	99.85 % (± 0.26 %)	98.06 % (± 2.13 %)
TruSeq mechanical fragmentation	100.00 % (± 0.00 %)	99.82 % (± 0.18 %)
	99.9 % (± 0.03 %)	99.96 % (± 0.07 %)
	100.00 % (± 0.02 %)	99.89 % (± 0.15 %)
Nextera DNA Flex enzymatic fragmentation -'on-bead tagmentation'	99.94 % (± 0.09 %)	99.72 % (± 0.18 %)
	99.95 % (± 0.05 %)	99.75 % (± 0.12 %)
	99.94 % (± 0.07 %)	99.74 % (± 0.15 %)

A 20x coverage was achieved by a high proportion of bases (> 99.8 %) independent of the technology used. Nevertheless, the TruSeq technology exhibited best performance (≥ 99.9 %) with the lowest deviations. Regarding the 50x coverage of core genes, the TruSight method resulted in apparent lower values (98.06 % compared to 99.89 % and 99.74 %). Best coverage was obtained using the TruSeq technology, followed by that of the Nextera Flex.

Gaps or rather low coverage regions have to be completed by Sanger sequencing in order to guarantee hundred percent coverage of core genes. For research applications, we set the coverage cut-off to ≥ 20x. On average, four exons in two genes had to be re-sequenced in each sample using the TruSight technology with strong deviations between samples. When applying the Nextera Flex protocol one exon in one gene had to be re-analyzed on average. The same exons, i.e. the first exon of the *KCNQ1* and

278 the fourth exon of the *KCNH2* gene were involved in multiple samples. Almost no target regions had to
279 be completed by using the TruSeq technology. Only few samples were involved and the same regions
280 had to be re-sequenced as using Nextera Flex.
281
282 **4. Discussion**
283 In the past years, next-generation sequencing technologies have developed into a promising tool for
284 examining rare and heterogeneous genetic disorders and investigating the genetic background in
285 victims of sudden unexpected death [3, 7, 16]. As Sanger sequencing-based testing is limited to the
286 analysis of well characterized, most prevalent genes, NGS enables the analysis of a broad range of
287 targets within short turnaround time and with reduced cost [17]. A wide variety of kits is available, both
288 for custom and predesigned panels. In order to select the best solution for the field of application,
289 various aspects need to be considered, such as the size of the region of interest, compatibility with NGS
290 technologies, DNA requirements, practicability and cost per sample [3, 12, 13].
291
292 Whole exome sequencing (WES), frequently followed by condition-specific filtering of resulting data,
293 prevalently represents the method of choice investigating rare inheritable disorders and SUD cohorts.
294 However, this approach is more attractive for high-throughput laboratories and often results in uneven
295 coverage across and between genes. Furthermore, major candidate genes may be covered with poorer
296 quality than by defined gene panel analysis. Moreover, the higher risk of detecting incidental variants
297 should not be underestimated [17–21]. Highly multiplexed PCR approaches for target enrichment may
298 result in lower overall sequencing performance as, for example, coverage uniformity due to non-uniform
299 amplification of target regions. In addition, variants in primer binding sites may cause preferential or
300 mono-allelic amplification and the compatibility with low integrity samples may be limited [12–14].
301 Considering these facts, we decided to apply Illumina sequencing platform-compatible hybridization
302 capture-based library preparation workflows including the predesigned TruSight Cardio panel, which
303 covers the most relevant genes. It has been shown that well-designed hybridization-based assays offer
304 superior performance analyzing larger target regions and result in better coverage uniformity [12, 13].
305
306 The majority of standard NGS workflows and protocols are developed and optimized for high-quality
307 samples frequently processed in diagnostic analyses. Since the integrity of the starting material can
308 critically affect the success of targeted NGS library preparation [12, 22–24], we performed quality
309 control of the starting material. Our results indicated the expected high variability in DNA integrity,
310 especially in postmortem samples. Therefore, assessment of the sample integrity was very useful in
311 selecting the best DNA extracts per case. Unexpectedly, FFPE tissue specimens were often easier to
312 process with older cases or than e.g. blood and tissue samples that were not suitably stored. It is
313 generally known, that preservation of genomic DNA in FFPE complicates its use in many downstream
314 applications due to e.g. degradation and cross-linking between proteins and DNA. However, possibly

315   well-prepared and processed FFPE specimens better preserved genomic DNA. Therefore, this option

316   may be kept in mind when processing challenging cases, as we frequently obtained (also in subsequent

317   experiments, data not shown) usable DNA extracts and subsequent sequencing results from FFPE

318   specimens.

319   This approach supplemented with further intermediate quality control steps considerably minimized the

320   risk of losing single libraries during the preparation workflow and producing unreliable sequencing data

321   due to poor sample quality saving time and costs when working with heterogeneous as forensic sample

322   material.

323

324   An important first step for all targeted capture-based NGS applications is the consistent fragmentation

325   of gDNA by mechanical or enzymatic means [25, 26]. In contrast to mechanical methods, the efficiency

326   of enzymatic DNA fragmentation is highly dependent on the enzyme to DNA input ratio [23]. Adaption

327   of DNA or enzyme amounts and/or incubation time may improve this critical step, but may result in

328   inconsistencies in fragment size distribution due to over-fragmentation of low integrity samples and

329   insufficient fragmentation of intact samples, respectively. This may have downstream effects on library

330   preparation process as well as on sequencing performance [23, 25]. Using mechanical shearing,

331   accurate fragmentation is less dependent on the integrity of the starting material, but requires more

332   time and specific equipment. Therefore, we initially used the TruSight (enzymatic) for intact and TruSeq

333   technology (mechanical fragmentation) for more degraded samples. The new Nextera methodology

334   uses magnetic bead-linked transposome complexes binding and fragmenting fixed amounts of DNA.

335   Using this technology, over-fragmentation should be avoided, because DNA fragments remain attached

336   to the beads following tagmentation promising compatibility with variable input types [23].

337

338   According to the diagnostic standards, important quality parameters for genetic analyses using NGS

339   include, among others, the average sequencing depth, uniformity of coverage, the percentage of target

340   regions sequenced with an informative read depth greater than or equal to 20 as well as the coverage

341   of core genes [10, 12, 27, 28]. Guidelines explicitly indicate that the reliability of the analysis should not

342   be compromised with the transition from Sanger sequencing to NGS [10]. Therefore, the coverage of

343   15 core genes associated with SCD and arrhythmogenic disorders was evaluated according to the

344   recommendations of current guidelines [10, 15, 29]. Low coverage regions have to be completed by

345   Sanger sequencing. Re-sequencing of too many uncovered targets have an impact on turnaround time

346   and cost [17]. For diagnostic germline genetic testing, higher minimum read depths (30-50x) can be

347   defined depending on the panel used, especially concerning the coverage of major candidate genes.

348   Therefore, we also focused on the 50x target coverage as evaluation criterion.

349   In overall comparison of the results, we received adequate sequencing performances with each of the

350   three library preparation technologies tested. The mean read depth as well as the mean coverage of

351   target regions (20x) over the entire panel was as high as published previously [18]. The TruSeq

352 technology resulted in the highest evenness of coverage across the entire panel, the highest target
353 read depth and good coverage of core genes with low deviations between samples. These results were
354 closely followed by the Nextera Flex methodology. We observed that the TruSight technology was more
355 dependent on sample quality as expected. Furthermore, this method resulted in apparently lower values
356 at 50x coverage of core genes requiring re-sequencing of several exons in different genes. In contrast,
357 few and mostly the same exons had to be re-sequenced using the Nextera Flex or the TruSeq workflow.
358 This predominantly concerned high GC-rich sequences in the potassium channel encoding genes
359 *KCNQ1* und *KCNH2*, which is a known problem [17].
360
361 The performance of library preparation technologies can also be measured by the ease of use and the
362 amount of DNA required [12]. The Nextera Flex workflow is by far the easiest, fastest and most flexible.
363 Usually, it is not even necessary to quantify and normalize gDNA inputs above 50 ng, because library
364 yields are normalized by saturation of the enrichment bead-linked transposome complexes (eBLT) [23].
365 However, with regard to heterogeneous sample material and to obtain comparable results, we always
366 quantified and defined the input amount per sample. The Nextera Flex technology enables flexible DNA
367 input amounts. In subsequent experiments (data not shown), we went down to 50 ng input with limited
368 low-integrity forensic samples and obtained good-quality and comparable sequencing data (table 1). In
369 contrast, the TruSeq protocol represents the most time consuming workflow. DNA shearing by
370 sonification requires additional working steps resulting in more total hands-on time and many
371 laboratories may not have access to a Covaris shearing system. Furthermore, we used the highest
372 amount of starting material applying this method, which may be problematic when processing very
373 limited or low concentrated samples. Concerning assay time, the TruSight technology is in the middle
374 range. Commercially available 'bundled solutions' containing reagents for one library preparation
375 workflow and subsequent sequencing make this approach attractive for single or occasional
376 applications investigating SUD cases or arrhythmogenic diseases.
377

378 **5. Conclusion**
379 In the present study, we tested and compared three Illumina library preparation technologies for
380 applicability with heterogeneous sample material and subsequent genetic screening using NGS. Our
381 results show that using the comparably fast and flexible Nextera Flex technology, high-quality
382 sequencing data were obtained with samples exhibiting varying levels of degradation.
383 The TruSeq library preparation method performed slightly better, but is more time consuming, requires
384 higher input amounts and also specific equipment. Processing challenging samples showing high levels
385 of degradation, this technology might represent the last possibility to analyze exceptional cases.
386 Finally, the TruSight technology resulted in appropriate sequencing outcomes analyzing relatively intact
387 gDNA samples, but, nevertheless, did not reach the coverage values of the other technologies tested.

388 However, this workflow might be a convenient approach for occasional investigations of genes with
389 known cardiac associations.
390 The technologies evaluated here are not restricted to genetic screening purposes in SUD cases and
391 arrhyhmogenic disorders and might be supportive for further forensic examinations, in which genetic
392 analyses are based on heterogeneous sample material.
393

394 **Conflicts of interest**
395 The authors declare no conflicts of interest.
396

397 **Acknowledgements**
398
399
400

REFERENCES

1. Ware JS, Roberts AM, Cook SA (2012) Next generation sequencing for clinical diagnostics and personalised medicine: implications for the next generation cardiologist. Heart 98: 276–281. https://doi.org/10.1136/heartjnl-2011-300742.

2. van Dijk EL, Auger H, Jaszczyszyn Y et al. (2014) Ten years of next-generation sequencing technology. Trends Genet 30: 418–426. https://doi.org/10.1016/j.tig.2014.07.001

3. Brion M, Sobrino B, Martinez M et al. (2015) Massive parallel sequencing applied to the molecular autopsy in sudden cardiac death in the young. Forensic Sci Int Genet 18: 160–170. https://doi.org/10.1016/j.fsigen.2015.07.010

4. Scheiper S, Ramos-Luis E, Blanco-Verea A et al. (2018) Sudden unexpected death in the young - Value of massive parallel sequencing in postmortem genetic analyses. Forensic Sci Int 293: 70–76. https://doi.org/10.1016/j.forsciint.2018.09.034

5. Hertz CL, Christiansen SL, Ferrero-Miliani L et al. (2016) Next-generation sequencing of 100 candidate genes in young victims of suspected sudden cardiac death with structural abnormalities of the heart. Int J Legal Med 130: 91–102. https://doi.org/10.1007/s00414-015-1261-8

6. Bagnall RD, Weintraub RG, Ingles J et al. (2016) A Prospective Study of Sudden Cardiac Death among Children and Young Adults. N Engl J Med 374: 2441–2452. https://doi.org/10.1056/NEJMoa1510687

7. Bezzina CR, Lahrouchi N, Priori SG (2015) Genetics of sudden cardiac death. Circulation Research 116: 1919–1936. https://doi.org/10.1161/CIRCRESAHA.116.304030

8. Magi S, Lariccia V, Maiolino M et al. (2017) Sudden cardiac death: focus on the genetics of channelopathies and cardiomyopathies. J Biomed Sci 24: 56. https://doi.org/10.1186/s12929-017-0364-6

9. Bagnall RD, Crompton DE, Semsarian C (2017) Genetic Basis of Sudden Unexpected Death in Epilepsy. Front Neurol 8: 348. https://doi.org/10.3389/fneur.2017.00348

10. Matthijs G, Souche E, Alders M et al. (2016) Guidelines for diagnostic next-generation sequencing. Eur J Hum Genet 24: 1515. https://doi.org/10.1038/ejhg.2016.63

11. Børsting C, Morling N (2015) Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet 18: 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002

12. Mamanova L, Coffey AJ, Scott CE et al. (2010) Target-enrichment strategies for next-generation sequencing. Nat Methods 7: 111–118. https://doi.org/10.1038/nmeth.1419

13. Samorodnitsky E, Jewell BM, Hagopian R et al. (2015) Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. Hum Mutat 36: 903–914. https://doi.org/10.1002/humu.22825

436   14. Ikegawa S, Mabuchi A, Ogawa M et al. (2002) Allele-specific PCR amplification due to sequence
437   identity between a PCR primer and an amplicon: is direct sequencing so reliable? Hum Genet
438   110: 606–608. https://doi.org/10.1007/s00439-002-0735-1

439   15. Priori SG, Blomström-Lundqvist C, Mazzanti A et al. (2015) 2015 ESC Guidelines for the
440   management of patients with ventricular arrhythmias and the prevention of sudden cardiac death:
441   The Task Force for the Management of Patients with Ventricular Arrhythmias and the Prevention
442   of Sudden Cardiac Death of the European Society of Cardiology (ESC). Endorsed by:
443   Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 36: 2793–
444   2867. https://doi.org/10.1093/eurheartj/ehv316

445   16. Bagnall RD, Semsarian C (2017) Role of the molecular autopsy in the investigation of sudden
446   cardiac death. Progress in Pediatric Cardiology 45: 17–23.
447   https://doi.org/10.1016/j.ppedcard.2017.02.005

448   17. Novelli V, Gambelli P, Memmi M et al. (2016) Challenges in Molecular Diagnostics of
449   Channelopathies in the Next-Generation Sequencing Era: Less Is More? Front Cardiovasc Med
450   3: 29. https://doi.org/10.3389/fcvm.2016.00029

451   18. Pua CJ, Bhalshankar J, Miao K et al. (2016) Development of a Comprehensive Sequencing
452   Assay for Inherited Cardiac Condition Genes. J Cardiovasc Transl Res 9: 3–11.
453   https://doi.org/10.1007/s12265-016-9673-5

454   19. Green RC, Berg JS, Grody WW et al. (2013) ACMG recommendations for reporting of incidental
455   findings in clinical exome and genome sequencing. Genet Med 15: 565–574.
456   https://doi.org/10.1038/gim.2013.73

457   20. Anil Patwardhan, Jason Harris, Nan Leng et al. (2015) Achieving high-sensitivity for clinical
458   applications using augmented exome sequencing. Genome Med 7: 1–15.
459   https://doi.org/10.1186/s13073-015-0197-4

460   21. Stiles MK, Wilde AAM, Abrams DJ et al. (2020) 2020 APHRS/HRS Expert Consensus Statement
461   on the Investigation of Decedents with Sudden Unexplained Death and Patients with Sudden
462   Cardiac Arrest, and of Their Families. Heart Rhythm. https://doi.org/10.1016/j.hrthm.2020.10.010

463   22. Nietsch R, Haas J, Lai A et al. (2016) The Role of Quality Control in Targeted Next-generation
464   Sequencing Library Preparation. Genomics, proteomics & bioinformatics 14: 200–206.
465   https://doi.org/10.1016/j.gpb.2016.04.007

466   23. Bruinsma S, Burgess J, Schlingman D et al. (2018) Bead-linked transposomes enable a
467   normalization-free workflow for NGS library preparation. BMC Genomics 19.
468   https://doi.org/10.1186/s12864-018-5096-9

469   24. Viering E, Molitor J, Friedmann L et al. (2017) Implementation of Automated Sample Quality
470   Control in Whole Exome Sequencing. JLS 11. https://doi.org/10.17265/1934-7391/2017.06.001

471   25.   Head SR, Komori HK, LaMere SA et al. (2014) Library construction for next-generation
472   sequencing: overviews and challenges. Biotechniques 56: 61-4, 66, 68, passim.
473   https://doi.org/10.2144/000114133
474   26.   Knierim E, Lucke B, Schwarz JM et al. (2011) Systematic comparison of three methods for
475   fragmentation of long-range PCR products for next generation sequencing. PLoS ONE
476   6: e28240. https://doi.org/10.1371/journal.pone.0028240
477   27.   Gargis AS, Kalman L, Berry MW et al. (2012) Assuring the quality of next-generation sequencing
478   in clinical laboratory practice. Nat Biotechnol 30: 1033–1036. https://doi.org/10.1038/nbt.2403.
479   28.   Mertes F, Elsharawy A, Sauer S et al. (2011) Targeted enrichment of genomic DNA regions for
480   next-generation sequencing. Brief Funct Genomics 10: 374–386.
481   https://doi.org/10.1093/bfgp/elr033
482   29.   Grassi S, Campuzano O, Coll M et al. (2020) Genetic variants of uncertain significance: How to
483   match scientific rigour and standard of proof in sudden cardiac death? Legal Medicine 45: 101712.
484   https://doi.org/10.1016/j.legalmed.2020.101712
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507

Supplement 1: Genes included in the TruSight Cardio Panel, Illumina (174 genes)

ABCC9	DPP6	KCNQ1	PTPN11	TRDN
ABCG5	DSC2	KLF10	RAF1	TRIM63
ABCG8	DSG2	KRAS	RANGRF	TRPM4
ACTA1	DSP	LAMA2	RBM20	TTN
ACTA2	DTNA	LAMA4	RYR1	TTR
ACTC1	EFEMP2	LAMP2	RYR2	TXNRD2
ACTN2	ELN	LDB3	SALL4	VCL
AKAP9	EMD	LDLR	SCN1B	ZBTB17
ALMS1	EYA4	LDLRAP1	SCN2B	ZHX3
ANK2	FBN1	LMF1	SCN3B	ZIC3
ANKRD1	FBN2	LMNA	SCN4B	
APOA4	FHL1	LPL	SCN5A	
APOA5	FHL2	LTBP2	SCO2	
APOB	FKRP	MAP2K1	SDHA	
APOC2	FKTN	MAP2K2	SEPN1	
APOE	FXN	MIB1	SGCB	
BAG3	GAA	MURC	SGCD	
BRAF	GATAD1	MYBPC3	SGCG	
CACNA1C	GCKR	MYH11	SHOC2	
CACNA2D1	GJA5	MYH6	SLC25A4	
CACNB2	GLA	MYH7	SLC2A10	
CALM1	GPD1L	MYL2	SMAD3	
CALR3	GPIHBP1	MYL3	SMAD4	
CASQ2	HADHA	MYLK	SNTA1	
CAV3	HCN4	MYLK2	SOS1	
CBL	HFE	MYO6	SREBF2	
CBS	HRAS	MYOZ2	TAZ	
CETP	HSPB8	MYPN	TBX20	
COL3A1	ILK	NEXN	TBX3	
COL5A1	JAG1	NKX2-5	TBX5	
COL5A2	JPH2	NODAL	TCAP	
COX15	JUP	NOTCH1	TGFB2	
CREB3L3	KCNA5	NPPA	TGFB3	
CRELD1	KCND3	NRAS	TGFBR1	
CRYAB	KCNE1	PCSK9	TGFBR2	
CSRP3	KCNE2	PDLIM3	TMEM43	
CTF1	KCNE3	PKP2	TMPO	
DES	KCNH2	PLN	TNNC1	
DMD	KCNJ2	PRDM16	TNNI3	
DNAJC19	KCNJ5	PRKAG2	TNNT2	
DOLK	KCNJ8	PRKAR1A	TPM1	

508

509 **Supplement 2:**

510 <u>**Table S1**</u>**: Overview of input amounts and protocol adaptions in comparison to the original protocols.**

Technology	Input	Pre-enrichment	Enrichment
TruSight	50 ng	• <u>Tagment genomic DNA</u> 10 µl TDE1 (Tagment DNA Enzyme TDE)	
TruSeq	DIN ≥6: 200 ng DIN <6: 300 ng	• <u>PCR nano program</u> Step 2: 12 cycles	• <u>Pool libraries</u> 500 ng per pre-enriched library for enrichment 12-plex enrichment with 6000 ng total library mass • <u>Amplify enriched library</u> Step 2: 13 cycles
Nextera Flex	DIN ≥5: 200 ng DIN <5: 300 ng	• FFPE recommendations in the protocol were applied • <u>Amplify Tagmented DNA program</u> Step 2: 4 min. denaturation Step 3: 30 s denaturation	• FFPE recommendations in the protocol were applied • <u>Pool libraries</u> 500 ng per pre-enriched library for enrichment 12-plex enrichment with 6000 ng total library mass • <u>Amplify enriched library program</u> Step 1: 60 s denaturation Step 2: 20 s denaturation, 12 cycles

511
512 <u>**Table S2**</u>**: Comparison of Illumina library preparation technologies TruSight, TruSeq and Nextera Flex for applicability with heterogeneous sample material.**
513 The assessment is based on the comparison of defined criteria between the technologies tested in this study. Thus, the evaluation must be considered in relation to the other
514 methods and is not based on overall performance of the technology.

Technology	Uniformity of coverage	Target coverage 50x	Coverage of core genes 50x		Assay Time	Equipment	Input*
TruSight	97.05 % (± 1.46 %)	97.21 % (± 2.73 %)	98.06 % (± 2.73 %)	😐	😐	😐	50 ng
TruSeq	99.11 % (± 0.15 %) ☺	99.72 % (± 0.25 %) ☺	99.89 % (± 0.15 %)	☺ ★	☹	☹	200-300 ng
Nextera Flex	98.9 % (± 0.65 %)	99.59 % (± 0.23 %)	99.74 % (± 0.15 %)	☺	☺	😐	flexible range (50-300 ng tested) ☺

515 *Values concerning input amounts refer to initial gDNA inputs successfully tested in this study and may differ from values stated in corresponding protocol