

## Next-generation sequencing of ultra-low copy samples: From clinical FFPE samples to single-cell sequencing

Tatiana Dupinay<sup>1</sup>, Agnès Nguyen<sup>2</sup>, Séverine Croze<sup>3</sup>, Fabienne Barbet<sup>3</sup>, Catherine Rey<sup>3</sup>,  
Patrick Mavingui<sup>2</sup>, Michel Pépin<sup>1</sup>, Joël Lachuer<sup>3</sup> and Catherine Legras-Lachuer<sup>2,3,\*</sup>

<sup>1</sup>Pathogènes Emergents et Rongeurs Sauvages, USC 1233, VetAgro-Sup-Campus Vétérinaire de Lyon Bât 6, porte 6 - 1021 avenue Claude Bourgelat, 69280 Marcy l'Etoile, France,

<sup>2</sup>Dynamique Microbienne et Transmission Virale, UMR CNRS 5557 - UCBL -

USC INRA 1193 - ENVL, Campus de la Doua, Bat. André LWOFF, 43 bd du 11 Novembre 1918, France, <sup>3</sup>Profilexpert, Faculté de Médecine et de Pharmacie, 3453 CNRS-US7 Inserm, 8 avenue Rockefeller, aile C2, 69373 Lyon Cedex 08, France

### ABSTRACT

Next generation sequencing (NGS) has become a powerful approach in the field of infectious diseases and has revolutionized clinical microbiology and virology. However, despite the advances in these technologies, the use of NGS remains challenging for complex samples, such as ultra-low copy samples, samples with poor integrity of nucleic acids, and single cells. NGS is currently being adapted to the analysis of such complex samples through the development and standardization of robust methodologies for enriching and amplifying low copy genomes, for restoring and amplifying degraded nucleic acids, and for developing instruments for nanotechnology miniaturization of sample preparation. We discuss these innovative methodologies and the substantial progress in sample preparation methods.

**KEYWORDS:** pathogen discovery, FFPE samples, single cell, NGS, sample preparation

### ABBREVIATIONS

BSL : BioSafety Level  
cccDNA : Covalently Closed Circular DNA

DECAL : Differential Expression of Customized Amplification Libraries  
DOP-PCR : Degenerate Oligonucleotide Primed-PCR  
DPC : DNA-Protein Cross  
EBOV : Ebola Virus  
FACS : Fluorescence-Activated Cell Sorting  
FFPE : Formalin-Fixed, Paraffin-Embedded  
HBV : Hepatitis B Virus  
HCV : Hepatitis C Virus  
HIV : Human Immunodeficiency Virus  
HTLV : Human T Leukemia Virus  
IVT : In Vitro Transcription  
LADS : Linear amplification for deep sequencing  
LASL : Linker-amplified shotgun libraries  
LCM : Laser Capture Microdissection  
LM-PCR : Ligation Mediated-PCR  
MARV : Marburg Virus  
MCC : Merkel Cell Carcinoma  
MCPyV : Merkel Cell Polyomavirus  
MDA : Multiple Displacement Amplification  
MGS : Microarray-based Genomic Selection  
MIP : Molecular Inversion Probes  
MLST : Multi Locus Sequence Typing  
NGS : Next Generation Sequencing  
PCR : Polymerase Chain Reaction  
PI : Propidium Iodide  
RCA : Rolling Circle Amplification

---

\*Corresponding author:  
catherine.lachuer@univ-lyon1.fr

RDA	: Representational Difference Analysis
RNA-seq	: RNA-sequencing
RSV	: Respiratory Syncytial Virus
RT-PCR	: Reverse Transcriptase- PCR
SCOTS	: Selective Capture Of Transcribed Sequences
SHS	: Solution Hybrid Selection
SISPA	: Sequence-Independent Single Primer Amplification
SPIA	: Single-Primer Isothermal Linear Amplification
SVG	: Single Virus Genomics
WGA	: Whole Genome Amplification
WNV	: West Nile Virus
WTA	: Whole Transcriptome Amplification

## INTRODUCTION

Next generation sequencing (NGS) has become a powerful approach in the field of infectious diseases and has revolutionized virology and microbiology studies, leading to many publications and new pathogen genome sequences. NGS is useful for viral and microbial metagenomic sequencing, particularly in the study of communities that populate an organism or ecosystem at any given time. This includes a “core” set of *commensal* microorganisms combined with agents associated with acute or persistent infections. It should be possible to detect infection before the emergence of symptoms, which will have significant implications for prevention and health-care delivery. NGS for broad-spectrum detection has been intensively applied for tracking the spread of pathogens worldwide in epidemiological surveillance studies [1], for analyzing microbial and viral community diversity in metagenomic analyses [2], for monitoring potentially more virulent genotypes and drug-resistant genotypes, detecting unknown pathogens and discovering novel viruses [3, 4], as well as for studying gene expression of infected host cells [5]. NGS is also a great epidemiologic tool to sequence full-length viral and bacterial genomes and generate draft genomes in a few hours. This was the case in the outbreak that sickened thousands and killed over 50 people in Germany in 2011; the *Escherichia Coli* isolate implicated was completely sequenced in less than 24 hours.

In microbiology, an important feature to consider is that microorganism culture is time-consuming

and labor-intensive, and 99% of infectious agents have not yet been cultivated. In addition, infectious agents can rapidly adapt to their hosts and environments and produce many variants and quasispecies. For these reasons, NGS is particularly useful to generate thousands of sequence reads to identify any microorganisms, variants, or quasispecies, without a priori knowledge on their nature, in a few hours and directly from clinical and culture-independent samples.

## Applications of NGS in clinical microbiology and virology

More than 20 NGS instruments are on the market [6]. Three technologies are commonly used: the Roche/454 technology (Roche Applied Sciences, Basel, Switzerland) with the 454 GS-FLX and the bench 454 GS-Junior sequencers; the Illumina/Solexa technology (Illumina Inc., San Diego, CA) with the Genome Analyser II (GAIIx), the HiSeq 2000 and 2500, and the bench MiSeq sequencers; and the Life Technology (Applied Biosystem, Foster city, CA) with the SOLiD 5500 XL W System, the Ion Proton Sequencer, and the bench Ion PGM sequencer (Ion Torrent). More recently, the single-molecule HeliScope technology was developed by Helicos (Helicos, Cambridge, MA) and the PacBio RS technology by Pacific Biosciences (Pacific Biosciences Inc., CA). All of them involve sample preparation, sequencing, and data acquisition, but differ in terms of DNA amplification processes, sequencing chemistry, and data acquisition [6, 7]. Therefore, each of these technologies has its own specificities in terms of read length, run-time, accuracy, error rates, multiplexing, instrument cost, and sample cost. SOLiD and HeliScope systems generate short reads (35-75 bp), which may be suitable for applications that require a very high throughput of sequences such as re-sequencing but that are less suitable for *de novo* sequencing and metagenomics. PacBio is one of the newest NGS technologies. Individual read length (860-1,100 bp) can be generated without an amplification step. The Roche 454 GS-FLX Titanium and the Illumina HiSeq 2500 technologies, which deliver long reads (up to 450 and 150 bp, respectively), are more suitable for *de novo* assembly, metagenomic analysis, and virus discovery. The GS FLX Plus

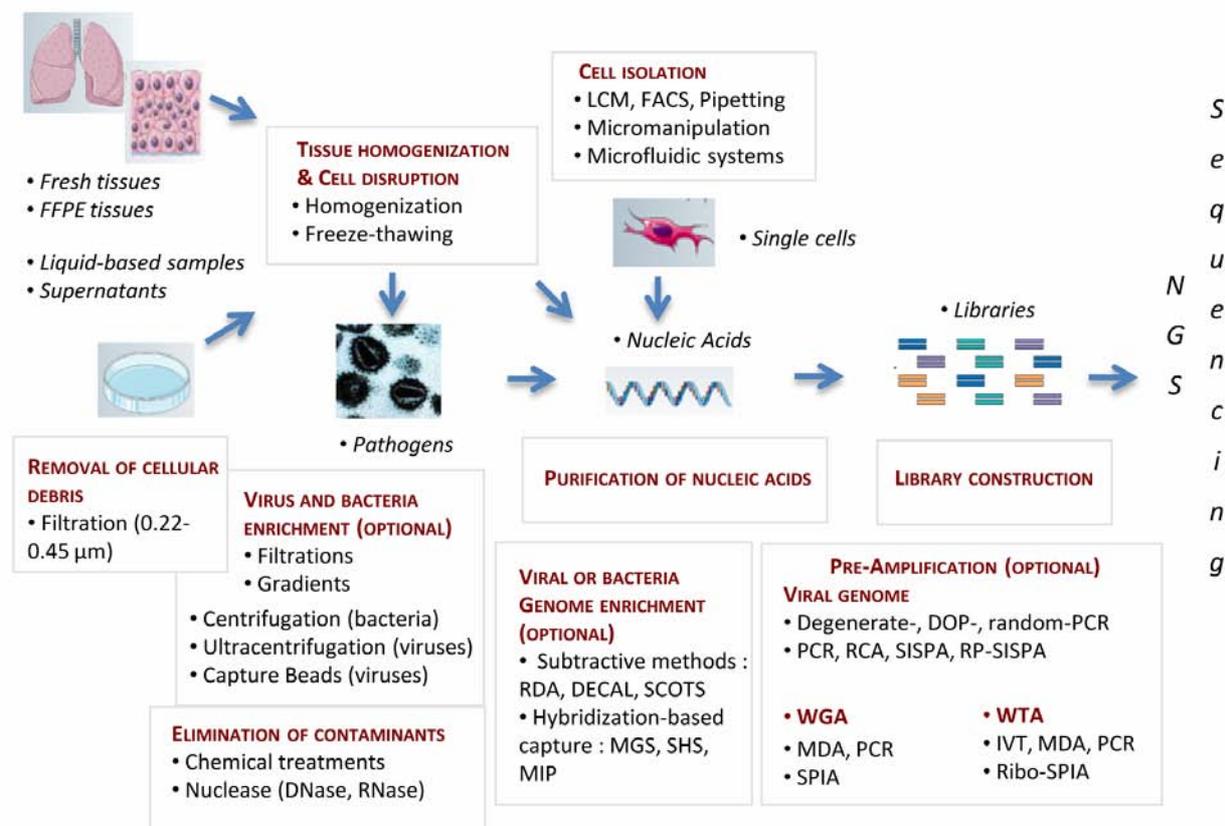
can produce up to one million reads per run of 700 bp and the Roche GS-Junior up to 100,000 reads of 400 bp. The Illumina HiSeq 2500 can produce 3 billion paired-end reads of 2\*150-bp and the MiSeq up to 30 million paired-end reads of 2\*250-bp.

Historically, Illumina GAIx and Roche 454 GS-FLX Titanium are the most commonly used platforms for microbial epidemiologic studies and pathogen discovery. Recent studies illustrate the successful use of Illumina GAIx for detection of H1N1 influenza A from nasopharyngeal swabs, at titers near the limits of detection by RT-PCR [8], and for the detection of antiviral therapy resistance mutations in the H1N1 influenza A neuraminidase gene at a sample fraction of 0.18% [9]. Recent comparisons between both Illumina GAIx and Roche 454 GS-FLX platforms, on various samples including the same microbial community DNA sample [10], a mixture of HIV clones [10, 11], H1N1 influenza A viral samples [12], and plasma samples from HIV-infected individuals [13] showed that despite the differences in read length, sequencing chemistries, and assembly strategies, the two platforms provide equivalent performances in terms of base-call errors, frame shift frequencies, and assemblies. However, rapid bench-top DNA sequencers such as the Ion Torrent PGM or the Illumina MiSeq have the advantages of producing sequences in less than a week (1 day vs ~2 weeks) and probably will be employed by users other than Genomics Centers and be integrated in routine clinical practices. For example, the Ion Torrent PGM has been recently used to determine the lineage of the *E. Coli* German outbreak strain [14]. The Illumina MiSeq has recently been evaluated in a pilot study of rapid bench-top sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance [15] and in a pilot study of rapid whole-genome sequencing for the investigation of a *Legionella* outbreak [16]. Both studies demonstrated the feasibility of using rapid whole genome sequencers to investigate outbreaks. These sequencers provide a powerful tool for clinical microbiology laboratories and possibly could become a routine monitoring tool for clinical diagnosis, epidemiology, and infectious disease surveillance.

However, despite the advances in NGS technologies, their routine use in clinical microbiology remains challenging. NGS is difficult both with low copy viral samples, such as clinical samples from patients receiving therapy, controller patients, and samples collected after peak viremia, and with diluted and large volume environmental samples (water, air). Routine use of NGS also remains challenging with regard to pathogen detection and identification from samples weakly represented, such as formalin-fixed, paraffin-embedded (FFPE) samples that contain nucleic acids of poor integrity, or infected tissue specimens in which the proportion of pathogen genome sequences is very low compared to the host genome sequences. Finally, analysis of infected sorted single cells, or uncultivable single bacterial cells, is promising but remains complex. NGS is currently being adapted to the analysis of such complex samples through the development and standardization of robust methodologies for the enrichment and amplification of low copy genomes at the single cell level, for the restoration and amplification of degraded nucleic acids, and for instrument development in nanotechnology miniaturization of sample preparation (Figure 1). These innovative methodologies and the substantial progress in sample preparation methods are discussed below.

### **Applications of NGS to low and ultra-low copy samples**

The first concern is that many of the studies reported above were performed using nucleic acids samples containing low levels of host nucleic acids, such as liquid-based samples (blood, serum, urine, nasopharyngeal swabs, cerebrospinal fluids, feces), culture supernatants, or bacterial isolates. Note that viral particle secretion in body fluids is time-limited to the period of viremia whereas detection in tissues could be more protracted, and therefore tissue sampling offers a better chance of identifying an offending pathogen and a broader picture of the viral components of a biome. However, NGS analysis of tissue samples requires considering several technical issues. The first is that the viral load within tissue samples is likely to be low. The second is the presence of host genome sequences from contaminating host cells, or naked DNA from disrupted cells, which can make up more



**Figure 1.** Flow chart of sample processing for NGS of pathogen samples.

than 99% of total nucleic acids. NGS alone can be insufficient if the viral or bacterial nucleic acids are in too low abundance relative to host nucleic acids. The number of viral reads obtained by deep sequencing from infected tissues is low, generally less than 0.1% of the total reads [17, 18]. Hence, it is necessary to eliminate host nucleic acids and to enrich bacterial or viral particles of interest. This is a three-stage process: (i) viral or bacterial concentration, (ii) removal of contaminating nucleic acids, and (iii) viral or bacterial genome enrichment, and amplification.

### Viral particle enrichment strategies

Several approaches have been developed for viral particle concentration. They are well documented and include various size selection filtrations, gradients, differential ultracentrifugation, and chemical and enzymatic pretreatments. The ultracentrifugation step is widely employed [19] but difficult to apply to a large series of samples.

A fast, simple, and reliable high-yielding method for viral particle recovery is tissue homogenization and cell disruption by freezing and thawing followed by filtering the samples through 0.22- and 0.45- $\mu\text{m}$ -pore-size discs. After homogenization of tissues, cells are disrupted by three freeze-thaw phases while leaving the nucleus intact. Nuclei are then pelleted by centrifugation and supernatants are treated by a cocktail of nucleases (RNase, DNase, Benzonase) to remove cellular nucleic acids and non-particle protected viral nucleic acids [20]. This method is based on the notion that the viral genome is protected within the nucleocapsid and capsid. Nuclease treatments need to be adapted to tissues and infectious agents and depend on the viral production itself. For viral metagenomic studies, microbial contamination can be reduced by filtering samples at 0.22- $\mu\text{m}$  to trap bacterial cells or by addition of solvents such as chloroform to permeabilize membranes of bacterial cells. Released chromosomal DNA is then

digested by DNase-1. Viral capsids are not sensitive to chloroform and remain intact.

Another important consideration is that sample processing methods can affect the composition of metagenomes. Although filtering at 0.22- $\mu\text{m}$  allows most viral particles to pass, large viral particles could be trapped on the filter, such as the giant Mimivirus, which is larger than some small bacteria, or EBV particles, which ranges in diameter from 120 to 220 nm. Recently, Willner *et al.* [21] showed that the relative abundances of phages in human oropharyngeal swabs change dramatically depending on whether samples are chloroform treated or filtered [21].

### Viral genome pre-amplification

Sequencing requires 1  $\mu\text{g}$  of DNA. Consequently, after purification, viral nucleic acid needs to be amplified to generate sufficient amounts of DNA for most sequencing platforms. RNA viruses have to be reverse-transcribed before amplification. Different approaches are available for viral genome amplification. Sequence-specific targeted viruses are widely employed; these generally use primers specifically designed to amplify specific RNA or DNA viruses. However, for viral metagenomics and virus discovery, viral genomes need to be amplified without prior viral sequence knowledge. Different sequence-independent methods have been developed: degenerate PCR, sequence-independent single primer amplification (SISPA), degenerate oligonucleotide primed (DOP)-PCR, random PCR, and rolling circle amplification (RCA) [22]. Three of them are more widely used: random PCR, SISPA, and RCA methods.

Random PCR for viral DNA and RNA library constructions uses two different primers: a first primer with a defined sequence at its 5' end, followed by a degenerate hexamer or heptamer sequence at the 3' end to randomly prime DNA synthesis, and a second primer complementary to the 5' defined region of the first primer [23] (Figure 2A). Following two rounds of extension to place primers at both extremities, multiple rounds of PCR amplification are performed with the defined sequence but lacking the degenerate 3' end. Random PCR is an established method for analyzing viromes [24], finding novel viruses [25], and detecting the presence of known viruses [24, 26].

The SISPA method, developed by Reyes and Kim [27], is a sequence-independent method based on ligation mediated (LM)-PCR. SISPA involves the partial cleavage of DNA by the Csp6.1 enzyme, followed by a directional ligation of an asymmetric adaptor to both ends of the DNA molecule (Figure 2B). Reverse-priming (RP)-SISPA adapted from SISPA was developed by Djikeng *et al.* [28, 29] to generate whole genome shotgun libraries of virus communities. In RP-SISPA, which is a combination of SISPA and random PCR, the cDNA is synthesized from RNA with a mixture containing a first primer with a 5' 20-bp sequence and a 3' random hexamer (N6) sequence and a second primer containing the same 5' 20-bp primer coupled with a 3' polyT tail. These SISPA and RP-SISPA amplification methods are widely used to characterize viruses from tissue samples and clinical biopsies [20, 30, 31] as well as for viral metagenome analyses [32]. Drawbacks of exponential based-PCR amplifications are the generation of bias such as the amplification of some regions more than others and the introduction of false-errors during polymerization.

The RCA method is an isothermal multiple displacement amplification (MDA) that uses phi29 DNA polymerase. RCA employs random hexamer primers that bind to multiple sites on the virus DNA genome and is based on the strong strand displacement activity of the phi29 DNA polymerase (Figure 2C). This polymerase has a good processivity and a low error rate (only 1 on  $10^6$ - $10^7$  bp). Viral DNA is exponentially amplified to generate micrograms of DNA [33]. But phi29 DNA polymerase cannot amplify RNA or short fragments such as cDNA. To overcome this, the method of Whole Transcriptome Amplification (WTA) has been combined with MDA. It includes a ligation step before the amplification, resulting in cDNA that are linked and then amplified by phi29 DNA polymerase [33-35]. Additionally, a combination of SISPA and RCA has been described [36]. One disadvantage of the MDA method, in particular for viral metagenomics, is the stochastic amplification bias, which makes the resulting metagenomes non-quantitative.

To circumvent bias resulting from PCR amplification, linker-amplified shotgun libraries (LASL) was

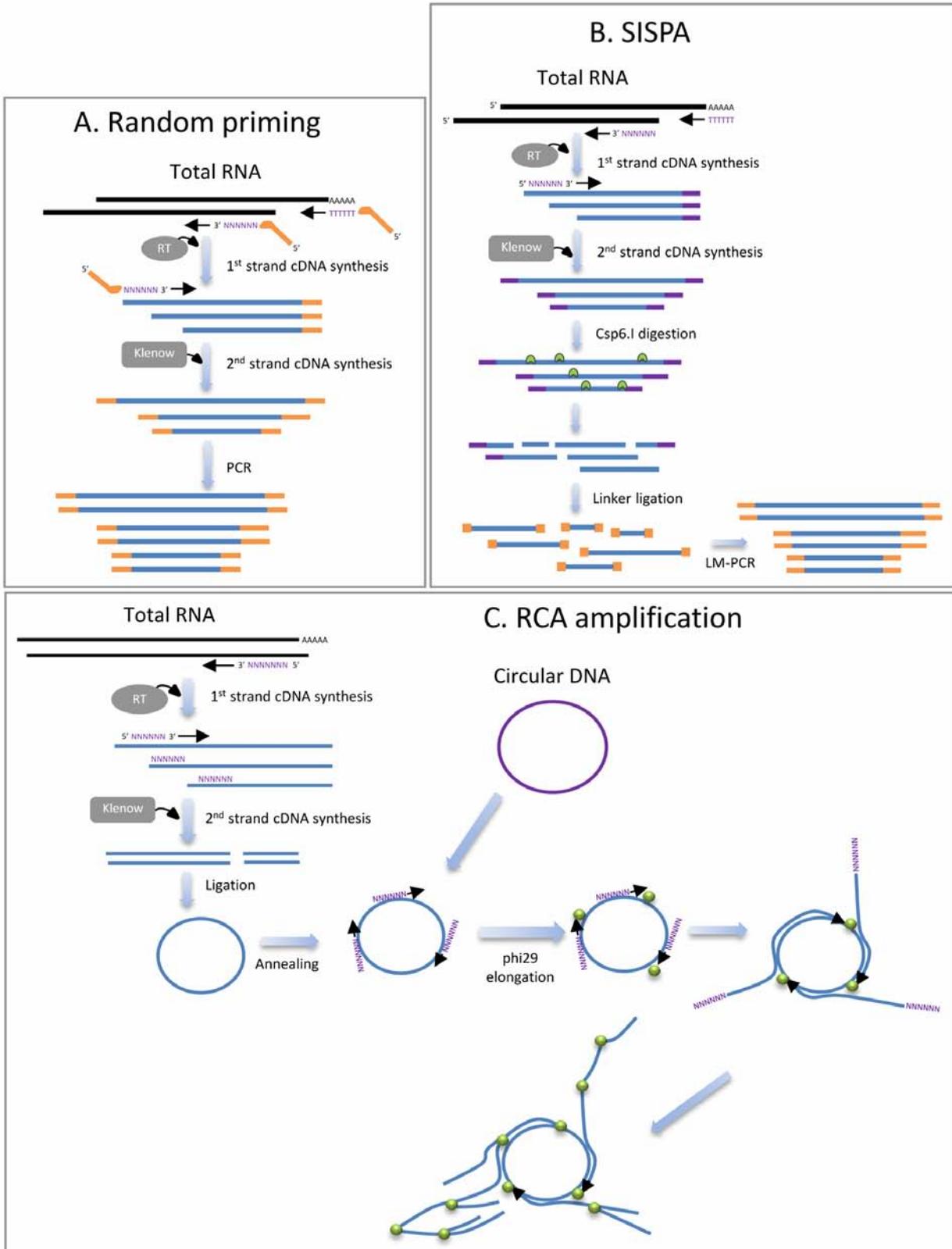


Figure 2

first described by Hoeijmakers *et al.*, [37] and then optimized by Duhaime *et al.* [38] for quantitative metagenomics of DNA viruses and other ultra-low DNA samples (as little as 1 pg of DNA). This method combines ligation mediated (LM)-amplification and *in vitro* transcription. It relies on attaching two different sequencing adapters to blunt-end repaired and a-tailed DNA fragments, wherein one of the adapters is extended with the sequence for the T7 RNA polymerase promoter. Ligated and size-selected DNA fragments are then transcribed *in vitro* and subsequent cDNA synthesis is initiated from a primer complementary to the first adapter.

There are number of commercially available whole genome amplification (WGA) methods designed to amplify extremely low quantities of DNA specifically for NGS platforms (Tables 1A and 1B) that have been used to amplify virus genomes. The *GenomePLEX DNA Amplification* kit from Sigma is based on PCR amplification. The *Illustra GenomiPhi V2 DNA Amplification* kit from GE HealthCare Life Sciences and the *Repli G* kit from Qiagen are based on MDA.

However, these methods may not be effective or applicable. For example, these strategies cannot be used to purify viruses that are in episomal forms, that do not produce viral particles, or that are integrated in the host genomes. Other approaches have been developed to overcome this, including total RNA sequencing (RNA-seq), and enrichment of infectious genomes by hybridization-based subtractive methods or capture-based methods.

### Total RNA sequencing

Another approach to virus discovery is to assume that infected cells produce viral transcripts, and that complexity can be reduced by sequencing total RNA without prior viral particle enrichment. Moore *et al.* [39] demonstrated by sequencing human RNA-seq libraries spiked with decreasing amounts of the Heterosigma *akashiwo* RNA-virus (HaRNAV) that the sensitivity of Illumina platform (GAIIx) is less than 1 in 1,000,000 reads, corresponding to 30,000 copies of viral RNA per sample. This method was used with input amounts of 500 pg to 100 ng of total RNA. And recently, Ninomiya *et al.* [40] demonstrated the ability of RNA-seq to differentiate hepatitis C virus (HCV) variants by sequencing total RNA purified from serum samples from patients with chronic HCV infection. For non-polyadenylated RNA, such as Dengue virus (DENV) RNA or bacterial RNA, ribosomal RNA (rRNA) depletion can be used to reduce complexity [41].

In case of ultra-low copy samples, the amplification of mRNA is necessary. Four major strategies of whole RNA amplification historically developed for microarrays and then adapted for single-cell transcriptome analysis, are currently available for RNA-seq of ultra-low copy samples. They include polymerase chain reaction (PCR)-based exponential amplification [42, 43], T7-based (IVT)-linear amplification [44, 45], a combination of PCR and IVT [46, 47], and single primer isothermal amplification (SPIA) developed by NuGEN [48, 49]. The SPIA method is a linear global RNA amplification method that uses a single chimeric

---

**Legend to Figure 2.** Random, SISPA and RCA amplification methods. A) Random amplification: Viral RNA (black bar) is converted into cDNA (blue bar) using random tagged primers and tagged polyT primers (purple letters and orange bar). After synthesis of second strand using the Klenow fragment of DNA polymerase in the presence of random primers, double stranded DNA is amplified by PCR. B) SISPA method: The SISPA method involves the transcription of viral RNA (black bar) to cDNA (blue bar) using random primers and polyT primers (purple letters). After the synthesis of second strand using the Klenow fragment of DNA polymerase, double stranded DNA is digested using the restriction enzyme Csp6.I followed by a directional ligation of asymmetric adaptors at both ends of the DNA molecule. Products are amplified by LM-PCR to generate whole genome shotgun libraries before sequencing. Adapted from Djikeng *et al.* (2008) [29]. C) RCA method: Viral RNA (black bar) is converted into cDNA (blue bar) using random primers and polyT primers (purple letters). The RCA method uses random hexamer primers that bind to multiple sites on a circular DNA or cDNA template and the phi29 polymerase (green circle) then exponentially amplifies small double stranded DNA (dsDNA) by multiple displacement amplification.

**Table 1A.** Low input DNA samples

Kits	<i>Illustra™ GenomiPhi™ V2 DNA Amplification Kit</i>	<i>Ampli WGA kit</i>	<i>ThruPLEX-FD kit</i>	<i>SeqPLEX amplification kit</i>	<i>Genome PLEX whole genome amplification (WG1 &amp; WG2) kit</i>	<i>Ovation WGA system kit</i>	<i>Repli G Midi/Mini kit</i>
Manufacturers/Suppliers	GE Healthcare Life Sciences	Silicone Genetics	Rubicon Genomics	Sigma Aldrich	NuGEN	Qiagen	
Web site	<a href="http://www.GElifesciences.com">www.GElifesciences.com</a>	<a href="http://www.siliconbio.com">www.siliconbio.com</a>	<a href="http://www.rubicongenomics.com">www.rubicongenomics.com</a>	<a href="http://www.sigmaaldrich.com">www.sigmaaldrich.com</a>	<a href="http://www.nugeninc.com">www.nugeninc.com</a>	<a href="http://www.qiagen.com">www.qiagen.com</a>	
Range input DNA	10 ng	10 pg	0.5 pg-50 ng	100 pg	10 ng	10 pg	10 pg
Based-technology	MDA	LM-PCR	LM-PCR	PCR	PCR	SPIA	MDA
Applications	Microarrays NGS	NGS	NGS (Illumina)	Microarrays NGS	Microarrays NGS	Microarrays	Microarrays NGS

**Table 1B.** Low input RNA samples.

Kits	<i>SMARTer cDNA synthesis low RNA kit</i>	<i>ExpressART mRNA amplification kit</i>	<i>Arcturus RiboAmp® HS PLUS Kit</i>	<i>Ambion® Message Amp™ II kit</i>	<i>TransPLEX Whole transcriptome amplification (C-WTA) kit</i>	<i>TransPLEX whole transcriptome amplification (WTA1) kit</i>	<i>Complete TransPLEX Whole Trans -criptome Amplification (WTA2) Kit</i>	<i>Ovation® One-Direct System kit</i>	<i>Quantitect whole transcriptome kit</i>
Manufacturers/Suppliers	Clontech	AmpTech	Life Technologies	Rubicon Genomics	Sigma Aldrich	NuGEN	Qiagen		
Web site	<a href="http://www.clontec.com">www.clontec.com</a>	<a href="http://www.amp-tec.com">www.amp-tec.com</a>	<a href="http://products.invitrogen.com">http://products.invitrogen.com</a>	<a href="http://www.rubicongenomics.com">www.rubicongenomics.com</a>	<a href="http://www.sigmaaldrich.com">www.sigmaaldrich.com</a>	<a href="http://www.nugeninc.com">www.nugeninc.com</a>			
Range input total RNA	100 pg	0.1ng to 700 ng	100 pg to 500 pg	1ng to 300 ng	20 pg to 2 ng	500 pg to 50 ng	5 to 25 ng	500 pg to 5 ng	20 ng
Based-technology	LD-PCR (long distance PCR)	IVT	IVT	PCR	PCR	Ribo-SPIA	PCR	Ribo-SPIA	MDA
Applications	NGS	Microarrays	Microarrays	Microarrays NGS	Microarrays NGS	Microarrays Beadarrays	Microarrays NGS	Microarrays RNA-seq	Microarrays

primer, a DNA polymerase with strand displacement activity, and RNase H for amplification of DNA (SPIA) and RNA (Ribo-SPIA) (Figure 3). To apply the RNA-seq method to ultra-low copy samples, Malboeuf *et al.* [50] evaluated single-primer isothermal linear amplification (SPIA), using NuGEN's Ovation RNA-seq system with viral samples, including human immunodeficiency virus (HIV), respiratory syncytial virus (RSV), and West Nile virus (WNV) samples and containing 500 ag to 240 fg, corresponding to 100 to 30,000 copies, of viral RNA. Coupling with the Illumina sequencing and optimized *de novo* assembly, they were able to generate full genome assemblies for HIV, RSV, and WNV with as little as 100 viral RNA genomes. Various commercial kits designed for whole RNA amplification are available, including the *TransPLEX WTA* kits from Sigma Aldrich, based on PCR amplification, the *Ambion Message Amp II* kit, based on IVT amplification, and the *Ovation pico WTA* kit from NuGEN, based on the Ribo-SPIA method (Table 1B).

#### **Viral or bacterial nucleic acid enrichment**

Hybridization-based subtractive methods such as Representational Difference Analysis (RDA) were first developed by Endoh *et al.* [51] and then successfully applied for viral nucleic acids enrichment and virus discovery [28]. However, the need to compare two samples, identical except for the presence or the absence, and the poor sensitivity and time-consuming nature of this method limit its use for clinical specimens and large-scale analyses. Subtractive methods have been developed for bacterial mRNA enrichment. They include the Differential Expression of Customized Amplification Libraries (DECAL) and the Selective Capture Of Transcribed Sequences (SCOTS) techniques that include PCR, subtractive hybridization, and selective capture of transcribed sequences steps. SCOTS was developed by Graham and Clark-Curtiss in 1999 [52] and used to identify expressed genes of *Mycobacterium tuberculosis* in response to phagocytosis by human macrophage cells [52], and infected mouse lungs in the chronic phases of infection [53], *Salmonella enterica* serovar Typhi [54], *Actinobacillus pleuropneumoniae* [55], and *Rickettsiales Ehrlichia ruminantium*, an intracellular ruminant bacteria [56].

#### **“Capture-by-hybridization” methods for target enrichment before NGS sequencing**

Several “capture-by-hybridization” strategies have been developed to enrich targeted sequences before NGS sequencing [57, 58]. They include solid phase hybridization with the Microarray-based Genomic Selection (MGS) method (Figure 4A), based on microarrays for capture [59], the Solution Hybrid Selection (SHS) method (Figure 4B), based on a solution hybrid selection using biotinylated probes that are then captured by streptavidin-coated magnetic beads [60], and the Molecular Inversion Probes (MIP) method (Figure 4C), based on single-stranded oligonucleotides that are converted into circular molecules in the presence of a specific target sequence [61]. Capture by SHS method generally involves the fragmentation of cDNA or genomic DNA followed by the construction of a library by ligation of adapters. Besides, hybridization generally requires more than 1 microgram of input DNA and therefore often requires pre-amplification. Pre-amplification is performed by LM-PCR with the same primers as for the library. Amplified DNA is hybridized to a complex mixture of capture biotinylated probes. Hybridized molecules are then heat-based eluted. After capture, the typical DNA yields from eluted samples are in the sub-microgram range and DNA needs to be amplified again before NGS, usually by LM-PCR. Commercial systems for capture-by-hybridization are available, including the *Sequence capture arrays* and the *SeqCapEZ* system from Roche NimbleGen, the *Agilent SureSelect* system, and the *FlexSelect* system from Flexgen. An alternative to oligonucleotides probes is the PCR-based enrichment method that uses PCR-derived capture probes to reduce the cost [62].

The suitability of the SHS method for enriching low amounts of viral nucleic acids was first demonstrated by directly sequencing 13 human herpesvirus genomes from a variety of clinical samples, including blood, saliva, vesicle fluid, cerebrospinal fluid, and tumor cell lines [63]. More recently, Singh *et al.* [62, 64] reported, as proof of concept, a novel PCR-based hairpin-primed multiplex amplification called MultiLocus Sequence Typing (MLST)-seq method that combines a PCR-based target enrichment method with NGS to enrich and discriminate closely

### D. Ribo-SPIA

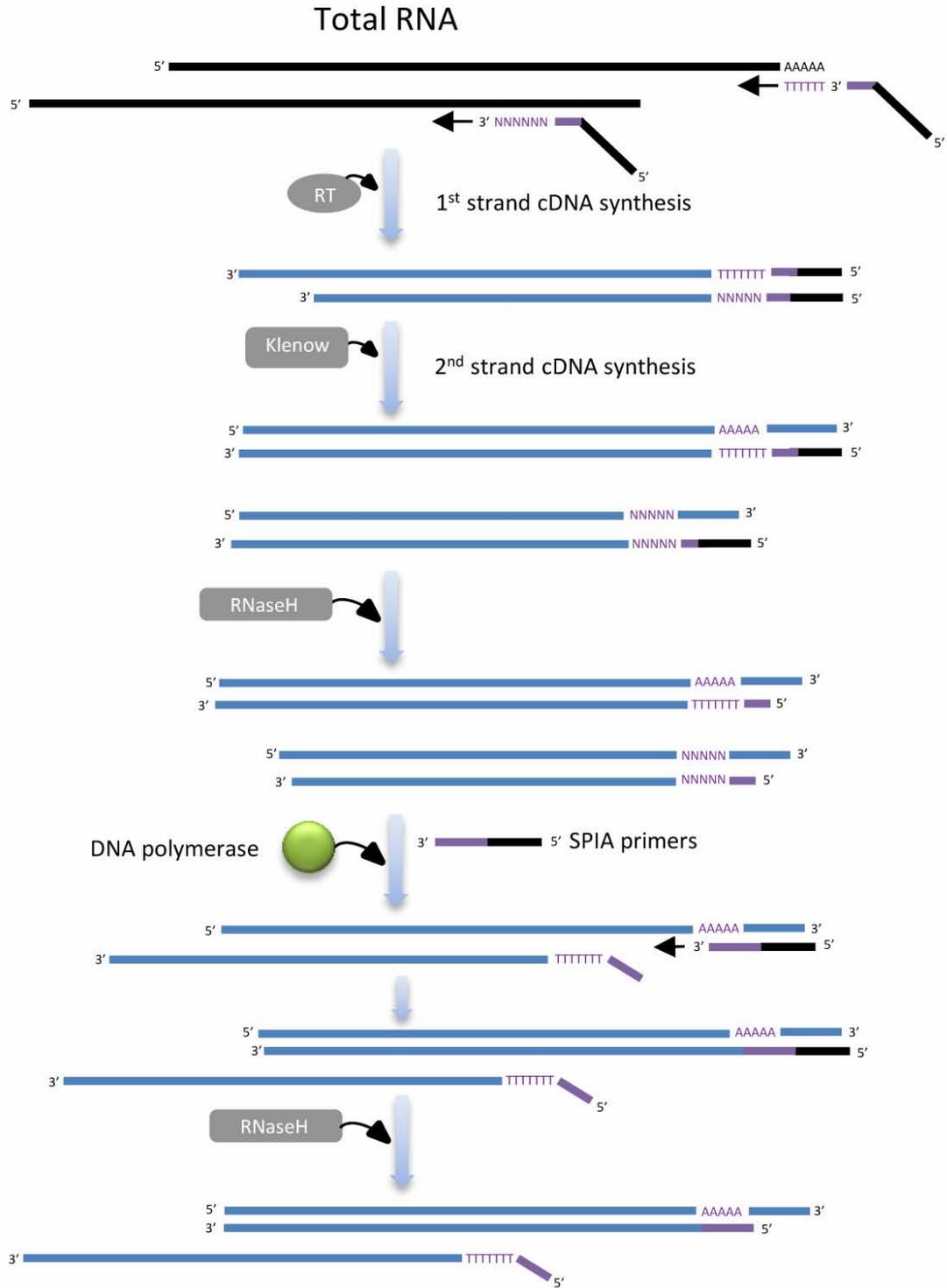


Figure 3

related strains of *Salmonella*. Based on the SHS method, our group has developed a capture system for blood-borne viruses (HIV-1, HIV-2, HTLV-1, HTLV-2, HCV, and HBV) that enhances the detection of these viruses by more than 800 times (unpublished data). Capture-hybridization enrichment strategies are also useful to enrich low-abundance cellular RNA. They have been successfully used to enrich virus encoded small non coding (snc) RNA present at a level of only 0.1 to 1% of all sncRNA in HIV-1 infected cells [65].

### Application of NGS to clinical samples of poor integrity

Another concern in clinical microbiology is the low copy nucleic acids in samples collected and stored into alcohol or formalin. This is an important component in pathogen surveillance and control of diseases. For screening of pathogens in vectors, sentinels, and reservoir populations, specimens must be collected on-site directly into 70%-80% alcohol or 10% formalin for transportation. Tissues containing infectious disease agents that require biosafety level (BSL)-3 and -4 need fixation times of 21 and 30 days, respectively. This concern is similar for FFPE clinical tissue. Because FFPE samples can be stored for several decades, they are an important resource for retrospective clinical studies. But FFPE specimens are challenging because the material is frequently degraded, and it contains substances such as formalin that could inhibit the linker ligation and the LM-PCR reaction during the library construction step. In addition, the

fixation induces chemical modifications of nucleic acids, DNA strand breaks, and modification of nucleotide such as DNA-protein cross-links (DPC). Hence, the detection of viral nucleic acids from FFPE tissues is considerably reduced. Using TaqMan assays from West Nile virus (WNV), Marburg virus (MARV), and Ebola virus (EBOV)-infected tissues, Mc Kinney *et al.* [66] reported a 2 log (10) reduction of detection with FFPE tissues compared to fresh tissues.

To date, and despite the poor efficiency of detection with FFPE tissues, many publications have reported on PCR, nested-PCR, and microarrays for the detection of viruses, such as the human papilloma virus (HPV) from cervical biopsies of intraepithelial neoplasia and squamous cell carcinoma [67], the hepatitis C virus (HCV) from liver biopsies [68], hepatitis B virus (HBV) covalently closed circular DNA (cccDNA) from small sections of liver biopsies [69], bacteria such as *Mycobacterium avium* [70], as well as for gene expression profiling. But there are still few usages of NGS for FFPE and fixed tissues. The need to enrich infectious agents is also crucial for FFPE samples, due to the severe DNA damage, the low input of DNA, and host genomic DNA interference, with the additional constraint that methods of particle enrichment and host genome depletion based on ultracentrifugation and nuclease treatment are difficult to apply to FFPE samples.

Enrichment based on laser capture microdissection (LCM) has been evaluated to enrich microbial fractions and applied to analyze the metagenomic profiles of *Helicobacter pylori* of archived formalin-fixed gastric section biopsies from two

---

**Legend to Figure 3.** Ribo-SPIA amplification method. Ribo-SPIA™ is an isothermal, linear global RNA amplification method that uses a RNA/DNA chimeric primer (black bar and purple letters) for the first strand cDNA synthesis (blue bar). The RNA template is then partially degraded in a heating step that also serves to denature the reverse transcriptase. DNA polymerase is added to the reaction mixture to carry out second strand cDNA synthesis forming a double stranded cDNA with a unique RNA/DNA heteroduplex at one end. This unique product serves as a substrate for the subsequent SPIA DNA amplification step. The amplification step is initiated by the addition of a reaction mixture containing a chimeric primer, a DNA polymerase with strong strand-displacement activity, and RNase H. The RNase H cleaves the RNA portion (black bar) of the heteroduplex at one end of the double-stranded cDNA, thus generating a unique partial duplex cDNA with a single-stranded DNA tail at the 3' end of the second-strand cDNA. This tail is the priming site for the SPIA amplification step. The sequence of the SPIA amplification primer is complementary to the sequence of the single-stranded 3' end of the second-strand cDNA in the partial duplex. DNA amplification is carried out by extension of this primer by a DNA polymerase with strand-displacement activity after cleavage of the RNA portion of the primer.

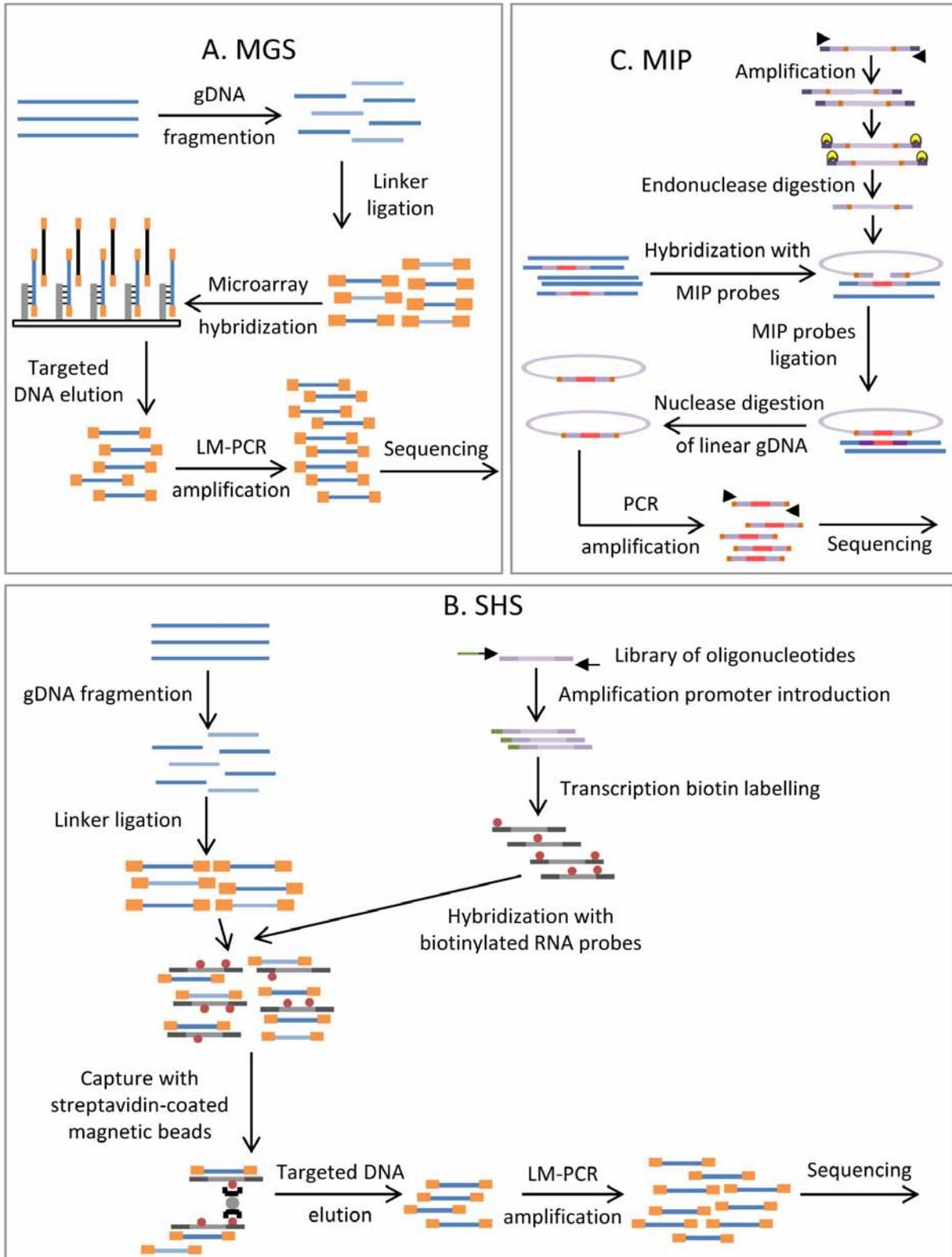


Figure 4

patients [71]. A whole genome pre-amplification step of 14 cycles was introduced previously to the library preparation for 454 sequencing [71]. About 5% of reads could be mapped to the *Helicobacter pylori* genome and 64 to 82.6% to the human genome. Considering the size of the human and the *H. pylori* genomes, these 5% reflect a proportion of about one contaminant human cell for more than 100 *H. pylori* cells, demonstrating the efficiency of enrichment. Recently, Conway *et al.* [72] reported on Illumina GAIIX coupled to LCM to investigate the presence of HPV in FFPE tissues from head and neck tumors. They showed that the amount of DNA extracted from LCM FFPE tumor specimens was sufficient to generate the libraries for sequencing. HPV sequences were detected in 10 of 31 samples with viral loads ranging from 1 to 98 HPV genomes per human genome and sensitivity was 50% compared to PCR and 75% compared to p16 immunohistochemistry. An alternative to LCM is the hybridization-based capture system described above. As proof of concept, Duncavage *et al.* [73] analyzed a 5.3-kb genome of Merkel cell polyomavirus (MCPyV) in cases of Merkel cell carcinoma (MCC). They generated PCR-derived capture probes, using biotinylated primers tiling across the MCPyV genome and demonstrated that this novel capture system coupled with Illumina GAIIX sequencing results in comparable efficiency in FFPE samples and fresh tissues. Pre-amplification of viral genomes

from FFPE samples may be done through commercially available WGA kits designed to amplify extremely small quantities or degraded/highly fragmented DNA (Table 2A) or RNA (Table 2B) specifically for NGS platforms.

### Single-cell sequencing

Yet another concern in clinical microbiology is the heterogeneity of tissues. There might be a low proportion of infected cells in the tissue and relatively few bacterial species might have been grown in pure culture, which requires sorting cells of interest. As a result, the possibility of sequencing genomes from single eukaryotic cell or single uncultivable bacterial cells is promising. In single-cell sequencing, cells are physically separated before sequencing and lysed to release nucleic acids, which are amplified and sequenced. Therefore, standard single-cell sequencing typically comprises four steps: (i) preparation of cells from samples, (ii) sorting and capture of cells for isolation of single cells, (iii) pre-amplification of nucleic acids, and (iv) library construction, sequencing, and data acquisition.

### Sorting and capturing cells for isolation of single cells

The prerequisites of methods for sorting and capturing cells for single-cell sequencing are (i) to separate and isolate specific targeted cells from a

---

**Legend to Figure 4.** Overview of capture-by-hybridization methods for target enrichment before NGS sequencing. Adapted from Summerer *et al.*, 2009 [58]. A) Microarray-based Genomic Selection (MGS) method: Variable length DNA probes were designed to tile across various segments of interest. Genomic DNA (blue bar) is randomly sheared, ligated to a linker at both ends, and hybridized to a capture array. Eluted products are amplified by LM-PCR before cluster generation and sequencing. B) Solution Hybrid Selection (SHS): A library of 200-mer oligonucleotides is synthesized. Each oligonucleotide consists of a target-specific 170-mer sequence (light purple) flanked by 15 bases of a universal primer sequence (purple) at both ends to allow PCR amplification. A T7 promoter (green) is added by PCR, and then *in vitro* transcription in the presence of biotin-UTP (red dot) generates a random single-stranded RNA capture probe library. In parallel, genomic DNA (blue bar) is randomly sheared and ligated to a linker on each side. Biotinylated probes and target genomic DNA are hybridized on a solution phase and the mix is incubated with streptavidin-coated magnetic beads to capture target-probe hybrids. Beads are washed. Hybridized fragments are eluted, amplified by LM-PCR, and analyzed using NGS. C) Molecular Inversion Probes (MIP): A library of DNA molecules is generated that contains a common internal linker sequence (light purple), two target-specific binding regions (purple), and two primer sequences (dark purple) containing an endonuclease recognition site. The library is amplified by LM-PCR and double digested with endonuclease restriction enzymes, resulting in a library of single-stranded DNA capture probes. Probes and target genomic DNA (blue bar) are hybridized and the targeted fragments (pink) are copied in a gap-filling reaction using DNA polymerase. The resulting target-probe hybrids are ligated by the DNA ligase, and non-circular probes are digested by an exonuclease. A library of circular target-probes is produced and amplified by LM-PCR before sequencing.

heterogeneous cell solution or tissue sample including other cells such as eukaryotic cells (epithelial, blood, immune cells) or commensal bacteria, (ii) to process samples within a short time to preserve cell viability and avoid nucleic acids degradation and (iii) to process samples, in a small volume (from nanoliters to femtoliters) to be compatible with the following steps (whole genome amplification, library construction, and NGS sequencing).

Methods available for single-cell isolation include serial dilutions, LCM [74], fluorescence-activated cell sorting (FACS) [75], manual picking using a mouth pipette [76], micromanipulation [77, 78] and microfluidic systems [79-81]. Manual picking, micromanipulation, and LCM have the advantages of allowing microscopic visualization and phenotyping of cells but they are time-consuming and more susceptible to contamination, and their uses are not applicable to high throughput analyses. FACS is a routine technology to isolate and sort eukaryotic cells and has been adapted for sorting prokaryotic cells [82, 83]. Fluorescent dyes such as SYTO specific dye incorporated by viable cells (green) and propidium iodide (PI) incorporated by non-viable cells (red) allow sorting of viable cells at the single-cell level [83]. The use of microfluidic systems to isolate and sort individual cells provides many advantages, including reduction of volumes and analysis time, enhancement of sensitivity, minimization of nucleic acid contamination from surrounding cells, and automation.

Many systems of microfluidics have been developed for sorting and capturing single cells, including methods based on selection adhesion to surface, cell migration inside a microfluidic device, dielectrophoresis, electrophoresis-based sequencing microchips, differential affinity, hydrodynamic trapping, optical and acoustic forces, gravity, magnetic, and droplet systems [84]. Droplets are well suited to the isolation of bacteria [85] and promising because multiple steps can be processed from cell sorting to the single-cell sequencing. Recently, a programmable droplet-based microfluidic device that combines the advantages of droplet-based sample compartmentalization (95 individual chambers) with the reconfigurable flow-routing control integrating microwaves technology has been developed for running bacterial cell sorting,

phenotyping, and single-cell whole genome amplification (WGA) [80]. This device has been successfully applied to diverse environmental samples, including marine enrichment culture, deep-sea sediments, and the human oral cavity [80].

Several companies offer miniature devices that integrate multiple steps for high-throughput analysis of single cells, such as the C1 Single-Cell Auto Prep System from Fluidigm (Fluidigm Corporation, San Francisco, CA), the DEParray from Silicon Biosystems, and the RainDance (RainDance Technologies, Lexington, MA). The C1 cell-sorting Single-Cell Auto Prep in combination with the BioMark™ HD System separates and captures up to 96 individual cells into individual chambers and processes single cells from lysis to gene expression profile for up to 96 transcripts. The cell sorting microarray platform (DEParray) is based on active microelectronic active silicon substrate embedding control circuitry for trapping single-cells in individual dielectrophoretic cages. It serves to detect, manipulate, and sort specific cells within a heterogeneous population, maintaining cell viability and integrity of nucleic acids.

After isolation of cells, the next steps are cell lysis to extract nucleic acids and amplification of extracted nucleic acids. The lysis treatments should be effective enough for sufficient yield but gentle enough not to damage nucleic acids and interfere with the following steps (reverse transcription, amplification). Chemicals such as phenol to remove proteins and membranes are not possible on single cells. Lysis can be achieved by chemical alkaline treatment (NaOH or KOH), physical treatment (heat, freeze-thawing), or enzymatic treatment (lysozyme, protease). Single-cell genome corresponds to a few femtograms (0.9 to 85 fg) of DNA for a typical bacterial genome and a few picograms of DNA (46 pg for a human cell) for a eukaryotic genome. Because NGS sequencing requires 1 µg of DNA, genomic DNA from single-cell genome template needs to be amplified. A method widely used to amplify DNA from single-cell template is MDA-based WGA amplification through the isothermal amplification by the phi29 DNA polymerase. Phi29 DNA polymerase has a good processivity (70,000 bases every time it binds), generates large fragments with an average length > 10 kb (typically

**Table 2A.** FFPE DNA samples.

Kits	Ampli1 WGA kit	ThruPLEX-FD kit	Seq Plex DNA amplification kit	Genome PLEX whole genome amplification (WGA1 & WGA2) kit	Ovation WGA FFPE system kit	Repli G FFPE kit
Manufacturers/Suppliers	Silicone Genetics	Rubicon Genomics	Sigma Aldrich	NuGEN	Qiagen	
Web site	<a href="http://www.siliconbiosystems.com">www.siliconbiosystems.com</a>	<a href="http://www.rubicongenomics.com">www.rubicongenomics.com</a>	<a href="http://www.sigmaldrich.com">www.sigmaldrich.com</a>	<a href="http://www.nugeninc.com">www.nugeninc.com</a>	<a href="http://www.qiagen.com">www.qiagen.com</a>	
Minimum input total DNA (cells)	10 pg (100 to 1000 cells)	10 pg	100 pg	10 ng	100 ng	100 to 300 ng
Based-technology	LM-PCR	LM-PCR	PCR	PCR	Ribo-SPIA	MDA
Applications	NGS	NGS (Illumina)	Microarrays NGS	Microarrays NGS	Microarrays NGS	NGS

**Table 2B.** FFPE RNA samples.

Kits	ExpressART FFPE amplification (nano) kit	ExpressART trinucleotide amplification kit For degraded RNA	Ovation FFPE WTA system kit	Ovation RNA-seq FFPE kit
Manufacturers/Suppliers	AmpTec		NuGEN	
Web site	<a href="http://www.amp-tec.com">www.amp-tec.com</a>		<a href="http://www.nugeninc.com">www.nugeninc.com</a>	
Range input total RNA	5 to 700 ng	0.1 ng to 300 ng	50 ng	100 ng
Based-technology	IVT	IVT	Ribo-SPIA	Ribo-SPIA
Applications	Microarrays	Microarrays	Microarrays	RNA-seq

10-40 bp in length) compared to 3 kb by Taq DNA polymerase, and can yield the femtograms to micrograms of DNA required for NGS. Historically, MDA was pioneered in 2002 by Lasken and colleagues [86] to amplify whole genome DNA from human cells, and then was adapted to amplify DNA from individual cells [87, 88]. Commercially available WGA and WTA kits have been designed to provide whole genome DNA and whole transcriptome amplification from single cells (Tables 3A, 3B). They are based on exponential or isothermal linear amplifications. Generally, they use lysates from single-cells, or cell pellets as input without further purification and include the library construction prior to sequencing. Commercially devices have been designed to produce libraries, such as the digital Mondrian microfluidic from NuGEN (NuGEN, San Carlos, CA) that is based on the spread out of droplet samples on hydrophobic surface by electrowetting. It produces libraries in 8 sample batch sizes suitable for the Illumina sequencing platform, and starting with as little as 0.1 to 1 ng of DNA.

Single-cell sequencing is especially well suited to study uncultivated microorganisms such as symbionts and was intensively used to obtain partial or complete genome sequences of various microorganisms. Kvist *et al.* [78] were the first to isolate a single *Archaea* cell from soil samples by

micromanipulation combined with fluorescence *in situ* hybridization. Isolated single cells served as template for MDA using phi29 DNA polymerase, and the MDA products were used for 16S rRNA gene sequencing. Single-cell genome sequencing was then applied to various microorganisms such as a candidate division OP11 [89], *E. coli* and other protists [88], bacterial symbionts [75], insect symbionts [77, 90], and vertebrate symbionts [79, 88] such as the intracellular filamentous bacteria (SFB) *Arthromitus Clostridiaceae*. The latter is a host-specific symbiont present in the lower intestine of many vertebrates that plays a role in immune response and host protection from intestinal pathogens [79].

In the field of metagenomics, single-cell sequencing combined with metagenomics has led to considerable progress in the knowledge of the ocean microbial community [91] and has enabled the discovery of various marine microorganisms [92].

### Single-virus genomics (SVG)

Virus isolation depends on cultivable virus-host systems. For uncultivated viruses, such as most bacteriophages, the isolation and complete genome sequencing of individual virus genomes represents a significant benefit. Recently, the team of Lastken [93] published a proof-of-principle

**Table 3A.** Single-cell DNA samples.

Kits	<i>Ampli1</i> WGA kit	<i>PicoPLEX</i> -WGA kit	<i>GenomePlex Single cell Whole genome amplification (WGA4)</i> kit	<i>Repli G single cell</i> kit
Manufacturers/Suppliers	Silicone Genetics	Rubicon Genomics	Sigma Aldrich	Qiagen
Web site	<a href="http://www.siliconbiosystems.com">www.siliconbiosystems.com</a>	<a href="http://www.rubicongenomics.com">www.rubicongenomics.com</a>	<a href="http://www.sigmaaldrich.com">www.sigmaaldrich.com</a>	<a href="http://www.qiagen.com">www.qiagen.com</a>
Minimum input total DNA (cells)	Single cell	Single cell	Single cell	Single cell
Based-technology	LM-PCR	LM-PCR	PCR	MDA
One tube procedure including lysis step	Yes	Yes	Yes	Yes
Applications	NGS	Microarrays NGS (Illumina)	Microarrays NGS	Microarrays NGS

**Table 3B.** Single-cell RNA samples.

Kits	STRT protocol	<i>Smarter ultra low RNA Kit</i>	<i>WT ovation one direct amplification system kit</i>	<i>TargetAmp kit</i>
Manufacturers/Suppliers	Home-made	Clontech	NuGEN/ partner Rubicon	Epicentre Biotechnologies
Web site		<a href="http://www.clontech.com">www.clontech.com</a>	<a href="http://www.nugeninc.com">www.nugeninc.com</a>	<a href="http://www.epibio.com">www.epibio.com</a>
Minimum input total RNA (cells)	Single cell	10 pg to 10 ng	Single cell	Single cell
Based-technology	PCR	LD-PCR	SPIA amplification	IVT
One tube procedure including lysis step	Yes	Yes	Yes	No
Single cell	Yes	Yes	Yes	Yes

study addressing Single Virus Genomics (SVG). They mixed two known *E. coli* bacteriophages, T4 and Lambda, and sorted them *via* flow cytometry onto microscope slides with distinct wells containing agarose beads. Viral particles were then individually embedded within the agarose droplets with an additional layer of agarose, and used as templates for *in situ* MDA and sequencing. Using this novel SVG method, they demonstrated the feasibility of reading a single complete lambda genome with an average depth of coverage of 437-fold. This new SVG method has the potential to revolutionize the discovery of new viruses in a wide variety of fields, including viral and microbial biology, epidemiology, and ecology.

### CONCLUSION AND PROSPECTS

Over the past several years, NGS technologies have exploded and the rapidly growing applications have revolutionized many fields of biology. They are very attractive for infectious diseases because they can identify any microorganism, without a priori knowledge on its nature, in a few hours and directly from clinical and culture-independent samples. For these reasons, NGS provides a powerful tool for clinical diagnosis, epidemiology, and surveillance and will probably be integrated in routine clinical practices. The sharply reduced costs of bench-top sequencers will probably extend their use and make NGS as routine as microarrays and PCR technologies.

Recent applications to ultra-low copy DNA samples, poor integrity DNA samples, and complex

heterogeneous samples underline the need for innovation in sample preparation, especially in the repair of DNA damages for FFPE samples and samples difficult to lyse, in enrichment of viral or bacterial sequences of interest for complex heterogeneous samples, in linear amplification of DNA to reduce bias inherent to PCR, and in multiplexing methods to enhance the sequencing capacity and reduce sequencing costs.

An important issue to consider is that the amplification of a single-copy microbial genome, single viral genome, or ultra-low copies of nucleic acids is highly susceptible to contaminations. Specific care is required, such as clean rooms, sterilized equipment, and UV-irradiated disposables and reagents. The use of microfluidics for reducing the volume to the nanoliter scale also helps to minimize nucleic acid contamination and to reduce bias in amplification reactions. In addition to specialized infrastructures (clean room), single-cell sequencing requires specific equipment (cell sorters, microfluidics, robotic liquid handlers, NGS sequencers) and specific skills in cell sorting, nucleic acids amplification, and NGS and data analysis. To address these challenges and make these innovating technologies more accessible to the French scientific community, we established a core facility specialized in the field of Single-Cell genomics, P2i-Profilexpert ([www.profilexpert.fr/fr/presentation/p2i.html](http://www.profilexpert.fr/fr/presentation/p2i.html)).

### CONFLICTS OF INTEREST

None identified.

## ACKNOWLEDGEMENTS

Manuscript preparation was supported by the Virosca grant funded by the Fondation Alliance Biosecure and the region Rhône-Alpes-Feder. Tatiana Dupinay has a post-doctoral position, supported by the Veterinary School, VETAGROSUP.

## REFERENCES

1. Chan, J. Z., Pallen, M. J., Oppenheim, B. and Constantinidou, C. 2012, *Nat. Biotechnol.*, 30, 1068.
2. Sibley, C. D., Peirano, G. and Church, D. L. 2012, *Infect. Genet. Evol.*, 12, 505.
3. Tang, P. and Chiu, C. 2010, *Future Microbiol.*, 5, 177.
4. Barzon, L., Lavezzo, E., Militello, V., Toppo, S. and Palu, G. 2011, *Int. J. Mol. Sci.*, 12, 7861.
5. Westermann, A. J., Gorski, S. A. and Vogel, J. 2012, *Nat. Rev. Microbiol.*, 10, 618.
6. Glenn, T. C. 2011, *Mol. Ecol. Resour.*, 11, 759.
7. Beerenwinkel, N., Gunthard, H. F., Roth, V. and Metzner, K. J. 2012, *Front. Microbiol.*, 3, 329.
8. Greninger, A. L., Chen, E. C., Sittler, T., Scheinerman, A., Roubinian, N., Yu, G., Kim, E., Pillai, D. R., Guyard, C., Mazzulli, T., Isa, P., Arias, C. F., Hackett, J., Schochetman, G., Miller, S., Tang, P. and Chiu, C. Y. 2010, *PLoS One*, 5, e13381.
9. Flaherty, P., Natsoulis, G., Muralidharan, O., Winters, M., Buenrostro, J., Bell, J., Brown, S., Holodniy, M., Zhang, N. and Ji, H. P. 2012, *Nucleic Acids Res.*, 40, e2.
10. Luo, C., Tsementzi, D., Kyrpides, N., Read, T. and Konstantinidis, K. T. 2012, *PLoS One*, 7, e30087.
11. Zagordi, O., Daumer, M., Beisel, C. and Beerenwinkel, N. 2012, *PLoS One*, 7, e47046.
12. Watson, S. J., Welkers, M. R., Depledge, D. P., Coulter, E., Breuer, J. M., de Jong, M. D. and Kellam, P. 2013, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 368, 20120205.
13. Archer, J., Weber, J., Henry, K., Winner, D., Gibson, R., Lee, L., Paxinos, E., Arts, E. J., Robertson, D. L., Mimms, L. and Quinones-Mateu, M. E. 2012, *PLoS One*, 7, e49602.
14. Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N. J., Hentschke, M., Chen, W., Pu, F., Peng, Y., Li, J., Xi, F., Li, S., Li, Y., Zhang, Z., Yang, X., Zhao, M., Wang, P., Guan, Y., Cen, Z., Zhao, X., Christner, M., Kobbe, R., Loos, S., Oh, J., Yang, L., Danchin, A., Gao, G. F., Song, Y., Yang, H., Wang, J., Xu, J., Pallen, M. J., Aepfelbacher, M. and Yang, R. 2011, *N. Engl. J. Med.*, 365, 718.
15. Eyre, D. W., Golubchik, T., Gordon, N. C., Bowden, R., Piazza, P., Batty, E. M., Ip, C. L., Wilson, D. J., Didelot, X., O'Connor, L., Lay, R., Buck, D., Kearns, A. M., Shaw, A., Paul, J., Wilcox, M. H., Donnelly, P. J., Peto, T. E., Walker, A. S. and Crook, D. W. 2012, *BMJ. Open.*, 2, e001124.
16. Reuter, S., Harrison, T. G., Koser, C. U., Ellington, M. J., Smith, G. P., Parkhill, J., Peacock, S. J., Bentley, S. D. and Torok, M. E. 2013, *BMJ. Open.*, 3, e002175.
17. Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., Conlan, S., Quan, P. L., Hui, J., Marshall, J., Simons, J. F., Egholm, M., Paddock, C. D., Shieh, W. J., Goldsmith, C. S., Zaki, S. R., Catton, M. and Lipkin, W. I. 2008, *N. Engl. J. Med.*, 358, 991.
18. Kuroda, M., Katano, H., Nakajima, N., Tobiume, M., Ainai, A., Sekizuka, T., Hasegawa, H., Tashiro, M., Sasaki, Y., Arakawa, Y., Hata, S., Watanabe, M. and Sata, T. 2010, *PLoS One*, 5, e10256.
19. Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L. and Rohwer, F. 2009, *Nat. Protoc.*, 4, 470.
20. Daly, G. M., Bexfield, N., Heaney, J., Stubbs, S., Mayer, A. P., Palser, A., Kellam, P., Drou, N., Caccamo, M., Tiley, L., Alexander, G. J., Bernal, W. and Heaney, J. L. 2011, *PLoS One*, 6, e28879.
21. Willner, D., Furlan, M., Schmieder, R., Grasis, J. A., Pride, D. T., Relman, D. A., Angly, F. E., McDole, T., Mariella, R. P. Jr., Rohwer, F. and Haynes, M. 2011, *Proc. Natl. Acad. Sci. USA*, 108, Suppl. 1, 4547.
22. Bexfield, N. and Kellam, P. 2011, *Vet. J.*, 190, 191.
23. Allander, T., Tammi, M. T., Eriksson, M., Bjerkner, A., Tiveljung-Lindell, A. and Andersson, B. 2005, *Proc. Natl. Acad. Sci. USA*, 102, 12891.

24. Lysholm, F., Wetterbom, A., Lindau, C., Darban, H., Bjerkner, A., Fahlander, K., Lindberg, A. M., Persson, B., Allander, T. and Andersson, B. 2012, *PLoS One*, 7, e30875.
25. Ng, T. F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B. S., Wommack, K. E. and Delwart, E. 2012, *J. Virol.*, 86, 12161.
26. Wylie, K. M., Mihindukulasuriya, K. A., Sodergren, E., Weinstock, G. M. and Storch, G. A. 2012, *PLoS One*, 7, e27735.
27. Reyes, G. R. and Kim, J. P. 1991, *Mol. Cell. Probes.*, 5, 473.
28. Djikeng, A. and Spiro, D. 2009, *Future Virol.*, 4, 47.
29. Djikeng, A., Halpin, R., Kuzmickas, R., Depasse, J., Feldblyum, J., Sengamalay, N., Afonso, C., Zhang, X., Anderson, N. G., Ghedin, E. and Spiro, D. J. 2008, *BMC. Genomics*, 9, 5.
30. Rosseel, T., Scheuch, M., Hoper, D., De Regge, N., Caij, A. B., Vandenbussche, F. and Van Borm, S. 2012, *PLoS One*, 7, e41967.
31. Victoria, J. G., Kapoor, A., Dupuis, K., Schnurr, D. P. and Delwart, E. L. 2008, *PLoS Pathog.*, 4, e1000163.
32. Victoria, J. G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S. and Delwart, E. 2009, *J. Virol.*, 83, 4642.
33. Erlandsson, L., Rosenstjerne, M. W., McLoughlin, K., Jaing, C. and Fomsgaard, A. 2011, *PLoS One*, 6, e22631.
34. Berthet, N., Reinhardt, A. K., Leclercq, I., van Ooyen, S., Batejat, C., Dickinson, P., Stamboliyska, R., Old, I. G., Kong, K. A., Dacheux, L., Bourhy, H., Kennedy, G. C., Korfhage, C., Cole, S. T. and Manuguerra, J. C. 2008, *BMC. Mol. Biol.*, 9, 77.
35. Cheval, J., Sauvage, V., Frangeul, L., Dacheux, L., Guigon, G., Dumey, N., Pariente, K., Rousseaux, C., Dorange, F., Berthet, N., Brisse, S., Moszer, I., Bourhy, H., Manuguerra, C. J., Lecuit, M., Burguiere, A., Caro, V. and Eloit, M. 2011, *J. Clin. Microbiol.*, 49, 3268.
36. Wyant, P. S., Strohmeier, S., Schafer, B., Krenz, B., Assuncao, I. P., Lima, G. S. and Jeske, H. 2012, *Virology*, 427, 151.
37. Hoeijmakers, W. A., Bartfai, R., Francoijs, K. J. and Stunnenberg, H. G. 2011, *Nat. Protoc.*, 6, 1026.
38. Duhaime, M. B., Deng, L., Poulos, B. T. and Sullivan, M. B. 2012, *Environ. Microbiol.*, 14, 2526.
39. Moore, R. A., Warren, R. L., Freeman, J. D., Gustavsen, J. A., Chenard, C., Friedman, J. M., Suttle, C. A., Zhao, Y. and Holt, R. A. 2011, *PLoS One*, 6, e19838.
40. Ninomiya, M., Ueno, Y., Funayama, R., Nagashima, T., Nishida, Y., Kondo, Y., Inoue, J., Kakazu, E., Kimura, O., Nakayama, K. and Shimosegawa, T. 2012, *J. Clin. Microbiol.*, 50, 857.
41. Bishop-Lilly, K. A., Turell, M. J., Willner, K. M., Butani, A., Nolan, N. M., Lentz, S. M., Akmal, A., Mateczun, A., Brahmabhatt, T. N., Sozhamannan, S., Whitehouse, C. A. and Read, T. D. 2010, *PLoS Negl. Trop. Dis.*, 4, e878.
42. Tietjen, I., Rihel, J. M., Cao, Y., Koentges, G., Zakhary, L. and Dulac, C. 2003, *Neuron*, 38, 161.
43. Gonzalez, J. M., Portillo, M. C. and Saiz-Jimenez, C. 2005, *Environ. Microbiol.*, 7, 1024.
44. Van Gelder, R. N., von Zastrow, M. E., Yool, A., Dement, W. C., Barchas, J. D. and Eberwine, J. H. 1990, *Proc. Natl. Acad. Sci. USA*, 87, 1663.
45. Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M. and Coleman, P. 1992, *Proc. Natl. Acad. Sci. USA*, 89, 3010.
46. Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K. D., Yamada, R. G., Ueda, H. R. and Saitou, M. 2006, *Nucleic Acids Res.*, 34, e42.
47. Kurimoto, K. and Saitou, M. 2011, *Methods Mol. Biol.*, 687, 91.
48. Dafforn, A., Chen, P., Deng, G., Herrler, M., Iglehart, D., Koritala, S., Lato, S., Pillarisetty, S., Purohit, R., Wang, M., Wang, S. and Kurn, N. 2004, *Biotechniques*, 37, 854.
49. Kurn, N., Chen, P., Heath, J. D., Kopf-Sill, A., Stephens, K. M. and Wang, S. 2005, *Clin. Chem.*, 51, 1973.

50. Malboeuf, C. M., Yang, X., Charlebois, P., Qu, J., Berlin, A. M., Casali, M., Pesko, K. N., Boutwell, C. L., Devincenzo, J. P., Ebel, G. D., Allen, T. M., Zody, M. C., Henn, M. R. and Levin, J. Z. 2013, *Nucleic Acids Res.*, 41, e13.
51. Endoh, D., Mizutani, T., Kirisawa, R., Maki, Y., Saito, H., Kon, Y., Morikawa, S. and Hayashi, M. 2005, *Nucleic Acids Res.*, 33, e65.
52. Graham, J. E. and Clark-Curtiss, J. E. 1999, *Proc. Natl. Acad. Sci. USA*, 96, 11554.
53. Azhikina, T., Skvortsov, T., Radaeva, T., Mardanov, A., Ravin, N., Apt, A. and Sverdlov, E. 2010, *Biotechniques*, 48, 139.
54. Daigle, F., Graham, J. E. and Curtiss, R. 3<sup>rd</sup>. 2001, *Mol. Microbiol.*, 41, 1211.
55. Baltes, N., Buettner, F. F. and Gerlach, G. F. 2007, *Vet. Microbiol.*, 123, 110.
56. Emboule, L., Daigle, F., Meyer, D. F., Mari, B., Pinarello, V., Sheikboudou, C., Magnone, V., Frutos, R., Viari, A., Barbry, P., Martinez, D., Lefrancois, T. and Vachiery, N. 2009, *BMC. Mol. Biol.*, 10, 111.
57. Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J. and Turner, D. J. 2010, *Nat. Methods*, 7, 111.
58. Summerer, D. 2009, *Genomics*, 94, 363.
59. Albert, T. J., Molla, M. N., Muzny, D. M., Nazareth, L., Wheeler, D., Song, X., Richmond, T. A., Middle, C. M., Rodesch, M. J., Packard, C. J., Weinstock, G. M. and Gibbs, R. A. 2007, *Nat. Methods*, 4, 903.
60. Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S. and Nusbaum, C. 2009, *Nat. Biotechnol.*, 27, 182.
61. Hardenbol, P., Baner, J., Jain, M., Nilsson, M., Namsaraev, E. A., Karlin-Neumann, G. A., Fakhrai-Rad, H., Ronaghi, M., Willis, T. D., Landegren, U. and Davis, R. W. 2003, *Nat. Biotechnol.*, 21, 673.
62. Singh, P., Foley, S. L., Nayak, R. and Kwon, Y. M. 2012, *J. Microbiol. Methods.*, 88, 127.
63. Depledge, D. P., Palser, A. L., Watson, S. J., Lai, I. Y., Gray, E. R., Grant, P., Kanda, R. K., Leproust, E., Kellam, P. and Breuer, J. 2011, *PLoS One*, 6, e27805.
64. Singh, P., Foley, S. L., Nayak, R. and Kwon, Y. M. 2013, *Mol. Cell. Probes*, 27, 80.
65. Althaus, C. F., Vongrad, V., Niederost, B., Joos, B., Di Giallonardo, F., Rieder, P., Pavlovic, J., Trkola, A., Gunthard, H. F., Metzner, K. J. and Fischer, M. 2012, *Retrovirology*, 9, 27.
66. McKinney, M. D., Moon, S. J., Kulesh, D. A., Larsen, T. and Schoepp, R. J. 2009, *J. Clin. Virol.*, 44, 39.
67. Morshed, K., Polz-Dacewicz, M., Szymanski, M. and Smolen, A. 2010, *Folia. Histochem. Cytobiol.*, 48, 398.
68. Dries, V., von Both, I., Muller, M., Gerken, G., Schirmacher, P., Odenthal, M., Bartenschlager, R., Drebber, U., Meyer zum Buschenfeld, K. H. and Dienes, H. P. 1999, *Hepatology*, 29, 223.
69. Zhong, Y., Han, J., Zou, Z., Liu, S., Tang, B., Ren, X., Li, X., Zhao, Y., Liu, Y., Zhou, D., Li, W., Zhao, J. and Xu, D. 2011, *Clin. Chim. Acta.*, 412, 1905.
70. Ikonopoulou, J., Gazouli, M., Pavlik, I., Bartos, M., Zacharatos, P., Xylouri, E., Papalambros, E. and Gorgoulis, V. 2004, *J. Microbiol. Methods.*, 56, 315.
71. Zheng, Z., Andersson, A. F., Ye, W., Nyren, O., Normark, S. and Engstrand, L. 2011, *PLoS One*, 6, e26442.
72. Conway, C., Chalkley, R., High, A., MacLennan, K., Berri, S., Chengot, P., Alsop, M., Egan, P., Morgan, J., Taylor, G. R., Chester, J., Sen, M., Rabbitts, P. and Wood, H. M. 2012, *J. Mol. Diagn.*, 14, 104.
73. Duncavage, E. J., Magrini, V., Becker, N., Armstrong, J. R., Demeter, R. T., Wylie, T., Abel, H. J. and Pfeifer, J. D. 2011, *J. Mol. Diagn.*, 13, 325.
74. Kang, Y., Norris, M. H., Zarzycki-Siek, J., Nierman, W. C., Donachie, S. P. and Hoang, T. T. 2011, *Genome Res.*, 21, 925.
75. Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., Dandekar, T. and Hentschel, U. 2011, *ISME J.*, 5, 61.
76. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K. and Surani, M. A. 2009, *Nat. Methods*, 6, 377.
77. Hongoh, Y., Sharma, V. K., Prakash, T., Noda, S., Taylor, T. D., Kudo, T., Sakaki, Y., Toyoda, A., Hattori, M. and Ohkuma, M. 2008, *Proc. Natl. Acad. Sci. USA*, 105, 5555.

78. Kvist, T., Ahring, B. K., Lasken, R. S. and Westermann, P. 2007, *Appl. Microbiol. Biotechnol.*, 74, 926.
79. Pamp, S. J., Harrington, E. D., Quake, S. R., Relman, D. A. and Blainey, P. C. 2012, *Genome Res.*, 22, 1107.
80. Leung, K., Zahn, H., Leaver, T., Konwar, K. M., Hanson, N. W., Page, A. P., Lo, C. C., Chain, P. S., Hallam, S. J. and Hansen, C. L. 2012, *Proc. Natl. Acad. Sci. USA*, 109, 7665.
81. Marcy, Y., Ouverney, C., Bik, E. M., Losekann, T., Ivanova, N., Martin, H. G., Szeto, E., Platt, D., Hugenholtz, P., Relman, D. A. and Quake, S. R. 2007, *Proc. Natl. Acad. Sci. USA*, 104, 11889.
82. Alvarez-Barrientos, A., Arroyo, J., Canton, R., Nombela, C. and Sanchez-Perez, M. 2000, *Clin. Microbiol. Rev.*, 13, 167.
83. Czechowska, K., Johnson, D. R. and van der Meer, J. R. 2008, *Curr. Opin. Microbiol.*, 11, 205.
84. Yin, H. and Marshall, D. 2012, *Curr. Opin. Biotechnol.*, 23, 110.
85. Zeng, Y., Novak, R., Shuga, J., Smith, M. T. and Mathies, R. A. 2010, *Anal. Chem.*, 82, 3183.
86. Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M. and Lasken, R. S. 2002, *Proc. Natl. Acad. Sci. USA*, 99, 5261.
87. Raghunathan, A., Ferguson, H. R. Jr., Bornarth, C. J., Song, W., Driscoll, M. and Lasken, R. S. 2005, *Appl. Environ. Microbiol.*, 71, 3342.
88. Marcy, Y., Ishoey, T., Lasken, R. S., Stockwell, T. B., Walenz, B. P., Halpern, A. L., Beeson, K. Y., Goldberg, S. M. and Quake, S. R. 2007, *PLoS Genet.*, 3, 1702.
89. Youssef, N. H., Blainey, P. C., Quake, S. R. and Elshahed, M. S. 2011, *Appl. Environ. Microbiol.*, 77, 7804.
90. Woyke, T., Tighe, D., Mavromatis, K., Clum, A., Copeland, A., Schackwitz, W., Lapidus, A., Wu, D., McCutcheon, J. P., McDonald, B. R., Moran, N. A., Bristow, J. and Cheng, J. F. 2010, *PLoS One*, 5, e10314.
91. Stepanauskas, R. 2012, *Curr. Opin. Microbiol.*, 15, 613.
92. Ishoey, T., Woyke, T., Stepanauskas, R., Novotny, M. and Lasken, R. S. 2008, *Curr. Opin. Microbiol.*, 11, 198.
93. Allen, L. Z., Ishoey, T., Novotny, M. A., McLean, J. S., Lasken, R. S. and Williamson, S. J. 2011, *PLoS One*, 6, e17722.