

Review

## Genome Walking by Next Generation Sequencing Approaches

Mariateresa Volpicella <sup>1,†</sup>, Claudia Leoni <sup>1,†</sup>, Alessandra Costanza <sup>1</sup>, Immacolata Fanizza <sup>1</sup>, Antonio Placido <sup>1</sup> and Luigi R. Ceci <sup>2,\*</sup>

<sup>1</sup> Department of Biosciences, Biotechnology and Pharmacological Sciences, University of Bari, Via Amendola 165/A, 70126 Bari, Italy; E-Mails: mariateresa.volpicella@biologia.uniba.it (M.V.); cla.leoni@biologia.uniba.it (C.L.); alessandra.costanza@biologia.uniba.it (A.C.); fanimma@libero.it (I.F.); antonio.placido@biologia.uniba.it (A.P.)

<sup>2</sup> Institute of Biomembranes and Bioenergetics, Italian National Research Council (CNR), Via Amendola 165/A, 70126 Bari, Italy

† These authors contributed equally to this work.

\* Author to whom correspondence should be addressed; E-Mail: l.ceci@ibbe.cnr.it; Tel.: +39-080-544-3311; Fax: +39-080-544-3317.

Received: 1 August 2012; in revised form: 31 August 2012 / Accepted: 25 September 2012 / Published: 1 October 2012

---

**Abstract:** Genome Walking (GW) comprises a number of PCR-based methods for the identification of nucleotide sequences flanking known regions. The different methods have been used for several purposes: from *de novo* sequencing, useful for the identification of unknown regions, to the characterization of insertion sites for viruses and transposons. In the latter cases Genome Walking methods have been recently boosted by coupling to Next Generation Sequencing technologies. This review will focus on the development of several protocols for the application of Next Generation Sequencing (NGS) technologies to GW, which have been developed in the course of analysis of insertional libraries. These analyses find broad application in protocols for functional genomics and gene therapy. Thanks to the application of NGS technologies, the original vision of GW as a procedure for walking along an unknown genome is now changing into the possibility of observing the parallel marching of hundreds of thousands of primers across the borders of inserted DNA molecules in host genomes.

**Keywords:** genome walking; insertional mutagenesis; next generation sequencing; functional genomics; gene therapy; transposon tagging; 454-pyrosequencing; Illumina; SOLiD; Ion Torrent

---

## 1. Introduction

The identification of unknown nucleotide sequences starting from a previously identified DNA region can be directly obtained by a number of Genome Walking (GW) methods all having in common a final PCR amplification in which an oligonucleotide specific for the known sequence is coupled with an oligonucleotide derived from the adopted GW strategy. The numerous GW methods can be classified into three main categories, according to the first step of the whole strategy: (1) **Restriction-based GW methods**, requiring a restriction digestion of genomic DNA and ligation of restriction fragments to DNA-cassettes; (2) **Primer-based GW methods**, in which PCR amplifications are directly carried out using a variously designed combinatorial (random or degenerate primer) coupled to a sequence specific primer; (3) **Extension-based GW methods**, in which the extension of a sequence specific primer and subsequent 3'-tailing of the resulting single strand DNA (ssDNA) provide the substrate for the final PCR amplification. Critical overviews of the conceived GW strategies, and their possible applications to both eukaryotic and prokaryotic genomes, have been recently reported [1–3].

GW is a highly flexible approach, allowing both the identification of specific, unique sequences (as for the analysis of single gene flanking sequences) and the analysis of large libraries (such as those obtained by insertional mutagenesis of retroviruses and transposable elements). In the latter case, GW has the potential to be powered by the enormous capacity of Next Generation Sequencing (NGS) approaches. In this review we focus on the application of NGS to GW. Among NGS technologies [4,5], currently only pyrosequencing and the sequencing by synthesis (SBS) methods have been reported as successful sequencing approaches for GW. A first report on the application of pyrosequencing (using the Roche 454 platform) to GW is that of Wang *et al.* [6], while the first application of SBS methods (using the Illumina technology) to GW is more recent [7]. Nowadays both technologies have been applied to GW by several research groups and, recently, specific methodological papers have been published [8–10]. In principle, however, also other NGS technologies in which parallel sequencing of amplicons is performed (such as the SOLiD “Sequencing by Oligo Ligation Detection” approach and the more recent “sequencing by synthesis” approach, known as Ion Torrent [11], by Life Technologies), can potentially be used for GW.

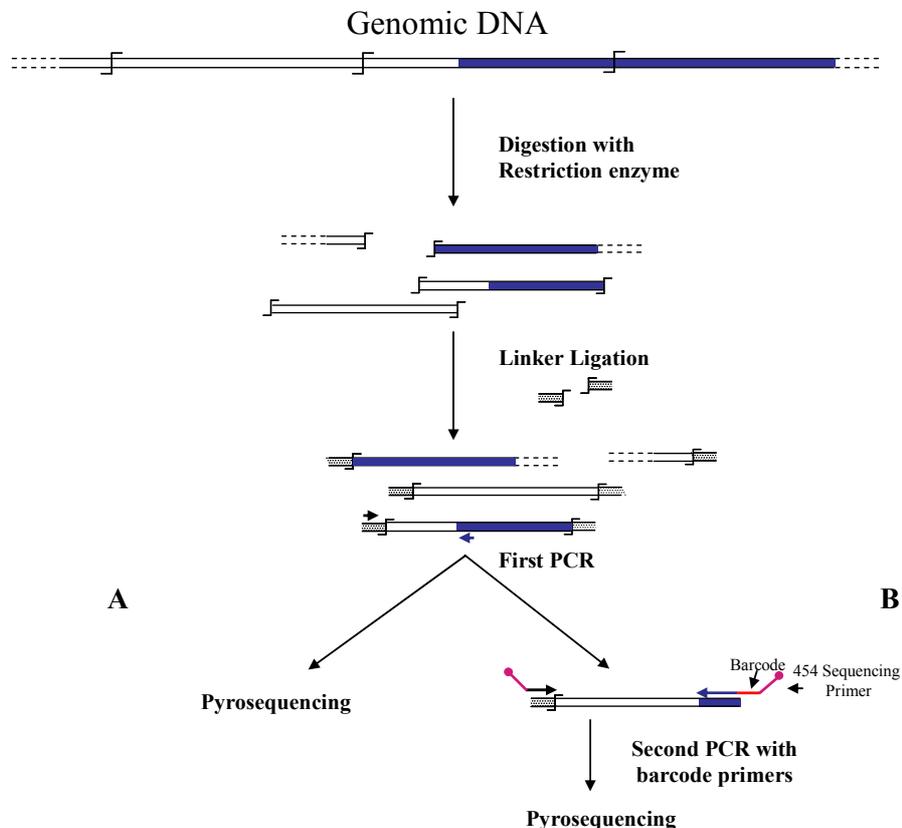
The first part of this review illustrates applications of pyrosequencing to GW, describing protocols adopted for the analysis of genomes of man, mouse, yeasts and plants, then the application of the SBS-Illumina technology to GW is reported. A final paragraph gives some general considerations about the possibility to use other, as yet unexplored, combinations of NGS and GW methods.

## 2. GW by 454-Pyrosequencing

Wang *et al.* [6] were the first to associate NGS and GW technologies to analyze the insertion sites of the immunodeficiency virus (HIV) into the human genome. In this first approach, authors used a

classical restriction-based GW approach, consisting essentially in the digestion of genomic DNA molecules with a restriction enzyme and subsequent ligation of restriction fragments to DNA cassettes. The library of PCR fragments obtained by amplifying the ligation products was directly used for pyrosequencing (Figure 1A). A similar approach, with the improvement of the introduction of biotinylated primers, but also with some complicated and unuseful steps, such as the ligation of a double set of adapters and a primer extension reaction to create an additional restriction site, was used to analyze the transposon flanking sequences in 1,000 petunia *dTph1* insertional mutants [12]. Interestingly, however, in this paper the use of a nucleotide barcode to distinguish specific samples was firstly adopted in NGS-GW experiments. At the same time, an improved version of their first NGS-GW approach, with the inclusion of a DNA barcoding step (Figure 1B) was also published by Wang *et al.* [13]. One hundred and sixty thousand integration sites for lentiviral and gamma-retroviral vectors in twenty-eight tissue samples from eight different mice could be identified by this approach.

**Figure 1.** Main steps of cassette-PCR genome walking (GW) and pyrosequencing association strategies. The original strategy reported by Wang *et al.* [6] (A) and its improvement by the adoption of barcoded primers [13] (B) are illustrated. Blue and white bars correspond to known and unknown sequences, respectively. A black horizontal arrow indicates adapter specific primers; a blue arrow represents primers designed in the known region. The  $\lrcorner$  symbol indicates a restriction site. Barcodes and 454 sequencing traits of longer amplification primers are indicated in the figure.



Liu *et al.* [14,15] combined another restriction-based GW approach, known as Digestion Ligation Amplification (DLA) with pyrosequencing. DLA-GW is characterized by the use of a single-stranded

oligonucleotide adaptor to ligate to restricted genomic DNA fragments. In this case the introduction of a DNA-barcode was also employed for sequencing in a single experiment multiple independent insertional mutants of the high copy number *Mu* transposon from different maize *Mu*-stocks. The assay allowed the observation of the expected *Mu*/genomic junctions in approximately 94% of the 965,000 reads, demonstrating the specificity of the strategy. 324 gene hotspots for *Mu* insertions were detected.

The first application of the pyrosequencing-GW approach to yeast is by Guo and Levin [16] who studied the integration of the *Tfl* retrotransposon in the genome of *Schizosaccharomyces pombe*. About 600,000 sequences were analyzed, allowing the identification of more than 73,000 independent integration sites.

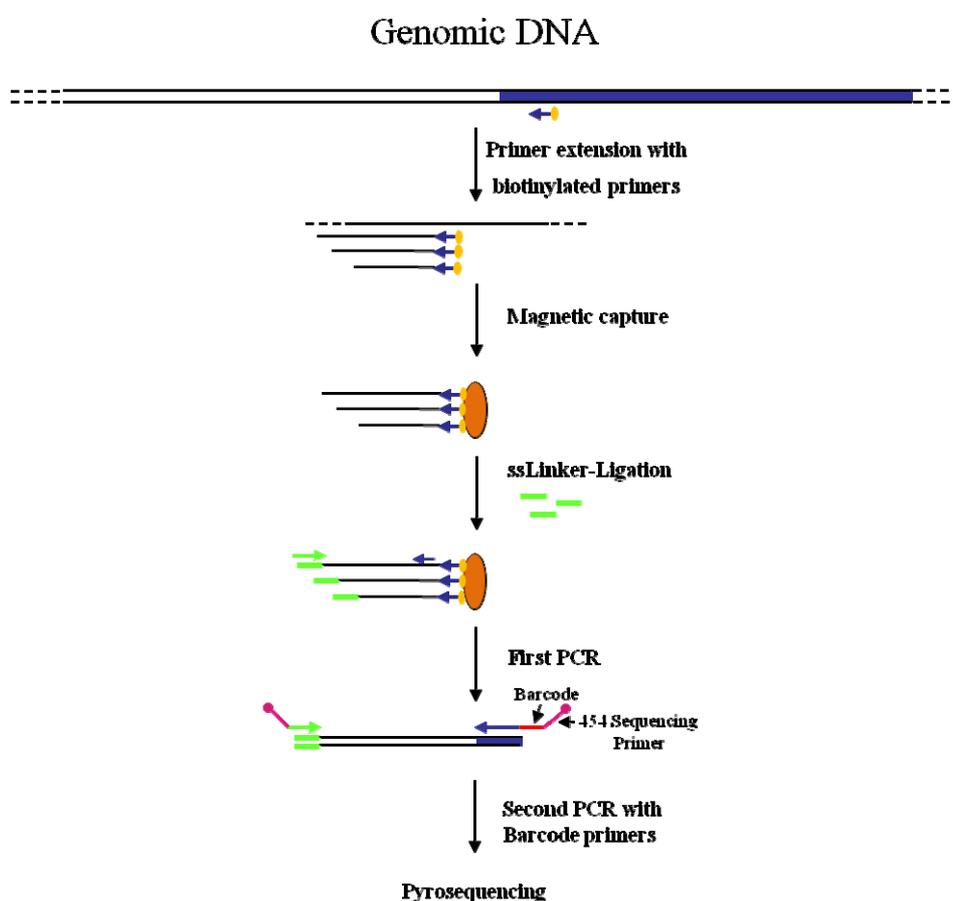
The above reported GW-pyrosequencing strategies adopted two restriction-based GW methods. Subsequently pyrosequencing was also applied to the Extension-based GW method known as nonrestrictive Linear Amplification-Mediated PCR (nrLAM-PCR) [9,10]. Extension-based GW methods are based on the extension of an oligonucleotide specific of the known DNA region and directed toward the unknown sequence. After the synthesis of the single-strand DNA, several strategies can be adopted to make it a suitable substrate for PCR reactions (reviewed in [1]). The basic method (known as Ligation Mediated PCR, LM PCR) was firstly introduced in 1989 by Muller and Wold [17,18] and several modifications, including the LAM-PCR, have been added since then. In the original LAM-PCR procedure [19] a 5'-biotynylated primer is extended from the known region of the genomic DNA and then a complementary strand is synthesized on the purified extension product by random exanucleotide priming. The obtained DNA molecule is then digested with a restriction enzyme recognizing a four nucleotide site, and ligated to a DNA cassette for providing the substrate for a final PCR amplification. In the “nonrestrictive” version of the method, the product of the extension reaction is ligated to a specific single-strand linker by means of RNA ligase. The strategy was designed in order to study the critical step of vector DNA integration during gene therapy, as a possible origin of the interruption of important genes and/or activation of proto-oncogenes by vector-introduced promoter and enhancer sequences. Figure 2 shows the combined nrLAM-pyrosequencing strategy, including barcoding of amplicons to facilitate clonal identification of the insertional events.

### 3. GW by Illumina-SBS

The first association of the SBS-Illumina NGS approach to GW was reported by Gawronski *et al.* [7] in their study for the identification of *Haemophilus influenzae* virulence genes required to delay bacterial clearance in the lungs of mice. The assay was carried out by sequencing insertion sites of a *Himar1-mariner* transposon insertional library of *H. influenzae* after infection in lungs of 5 mice (negative selection). Transposon/bacterial DNA junctions were identified in the resulting libraries by the so-called “high-throughput insertion tracking by deep sequencing” (HITS) method, practically consisting in a massive GW analysis by the Illumina technology (Figure 3). In this strategy the fragmentation of genomic DNA by restriction enzymes (as shown in Figure 1) has been replaced by “shearing”. The additional step of repairing DNA ends is however necessary before ligation of adapters. After repairing of the sheared DNA ends from transposon mutant libraries and ligation of Illumina oligonucleotide adapters, transposon-containing fragments are enriched via PCR using a biotinylated transposon primer and affinity purification on streptavidin-coated paramagnetic beads.

Purified fragments are then used for sequencing according to Illumina protocols. Using this approach, a large library of approximately 75,000 insertional mutants was analyzed, providing a rapid genome-wide analysis of bacterial genes required for growth/survival during infection of host organisms.

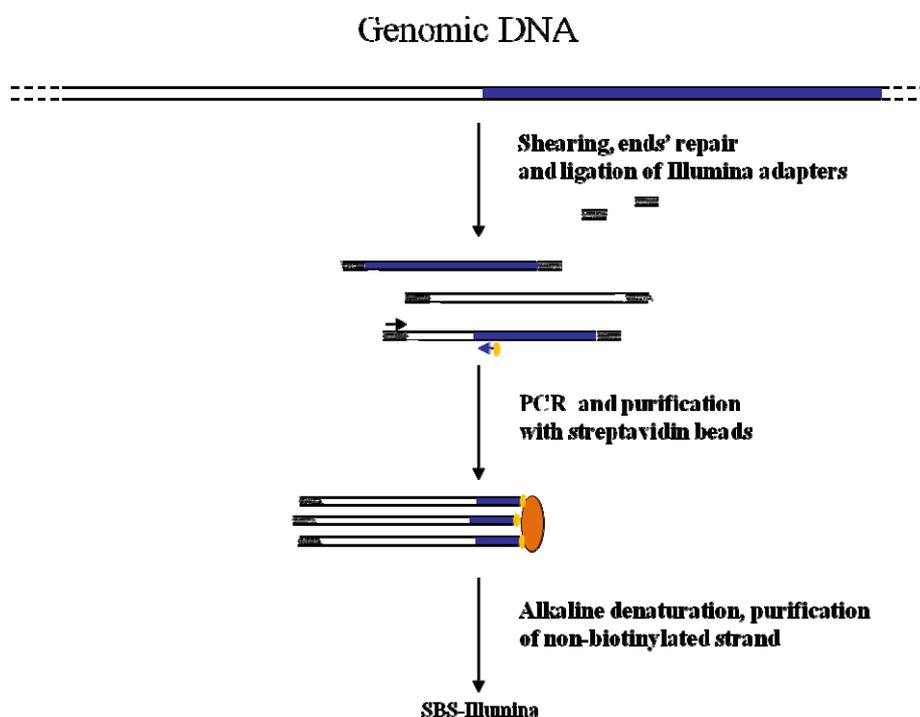
**Figure 2.** Main steps of nrLAM-PCR GW and pyrosequencing association strategy. Blue and white bars correspond to known and unknown sequences, respectively. A blue arrow with a small yellow oval corresponds to biotinylated primers in the known nucleotide region. An orange oval indicates a streptavidin magnetic device. Green and blue arrows represent primers designed in correspondence of the linker and known genomic region, respectively. Barcodes and 454 sequencing traits of longer amplification primers are indicated in the figure.



In an almost contemporary paper, a similar approach was used to investigate a large *Tn5*-derived bacterial transposon insertional library produced in *Salmonella enterica* for the identification of interrupted genes [20]. Sequencing was directly carried out on amplicons obtained with one primer corresponding to the transposon, and a second primer corresponding to the Illumina adaptor. Even if not mentioned by the authors, this approach stands as a classical NGS-GW approach. A *Tn5*-derived transposon was used to generate an estimated pool of 1.1 million transposon mutants and 370,000 unique transposon insertion sites were identified. Authors performed also a negative selection in order to identify *S. enterica* genes involved in the bacterium resistance to high concentrations of bile. 169 genes involved in bile tolerance were identified by this approach.

The same approach was successful for the analysis of an eukaryotic transposon insertional library. Li *et al.* [21] analyzed an inducible *piggyBac* (*PB*) transposon-based mutagenesis library in the yeast *S. pombe*. From a mutant pool of 400,000 Arg<sup>+</sup> Ura<sup>+</sup> colonies, *PB* insertions were detected in 54% of the ORFs containing the typical TTAA *PB* insertion site.

**Figure 3.** Main steps of cassette-PCR GW and sequencing by synthesis (SBS)-Illumina association strategy. Blue and white bars correspond to known and unknown sequences, respectively. A blue arrow with a small yellow oval corresponds to biotinylated primers in the known nucleotide region. A black horizontal arrow indicates adapter specific primers. An orange oval indicates a streptavidin magnetic device.



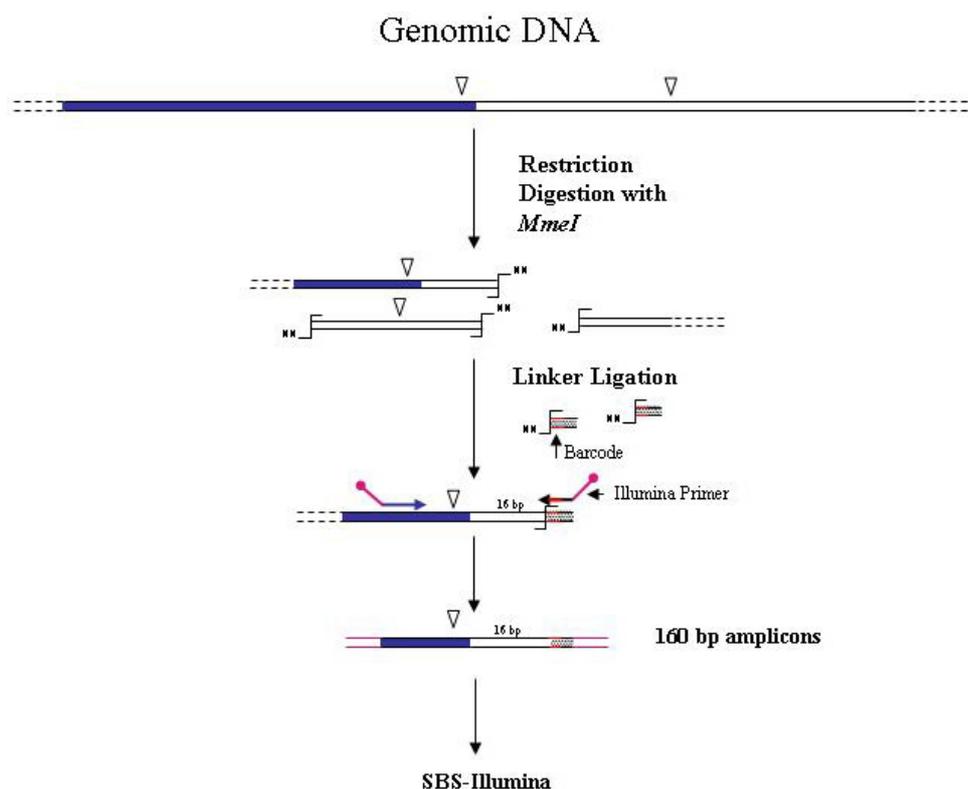
Illumina-GW proved also effective in the identification of *Mu* transposon insertions in the genome of maize photosynthesis mutants [22]. Distinct DNA samples were pooled and analyzed thanks to the introduction of barcodes. The problem of unambiguous assignment of transposon flanking regions, which may be encountered in the analysis of complex eukaryotic genomes, was addressed by authors by using 60-mer long biotinylated primer and low-cycle PCR amplifications. The approach allowed the identification of four genes whose interruption blocks chloroplast biogenesis. In a similar approach Urbanski *et al.* [23] analyzed a pool of Lotus mutant plants containing the retrotransposon *LORE1*. In this assay, the blunt-ended sheared DNA fragments were first subjected to 3' adenylation by *Taq* polymerase, and then ligated to splinkerette-adaptors provided of single T-overhangs. Thanks to this strategy 3,744 plants were examined and 8,935 new *LORE1* insertion sites were identified.

A different strategy for sequencing transposon/genomic DNA junctions was ideated by van Opijnen *et al.* [8,24] to identify the insertion sites of the *mariner* *Himar1* mini-transposon into the genome of the Gram-positive bacterium *Streptococcus pneumoniae*. The strategy is highly innovative and must not be confused with a restriction-based/NGS approach. Indeed in this case, differing from all the restriction-based GW methods, the enzyme recognition site is not on the genomic DNA of the mutated

organism, but in the transposon used for mutagenesis. The method relies on the presence of a *MmeI* recognition site at four bp from the left and right ends of the inverted repeats of the transposon. *MmeI* cuts 20 bp downstream of the recognition site and generates a random two-nucleotide 3'-overhang. After restriction of the library DNA, the ligation of an adapter with a random two-nucleotide overhang allows the amplification of unique-size fragments using transposon and adapter specific primers. Amplicons of 160 bp will contain 16 bp of flanking genomic DNA. The introduction of a DNA barcode was also employed for sequencing different samples in a single flow cell lane using the Illumina approach (Figure 4). The *S. pneumoniae* library was analyzed to categorize genes according to their relevance in bacterial fitness, allowing identification of genes essential for basal growth. A similar approach was also used by Goodman *et al.* [25] for the identification of fitness genes of the gut symbiont *Bacteroides thetaiotaomicron*.

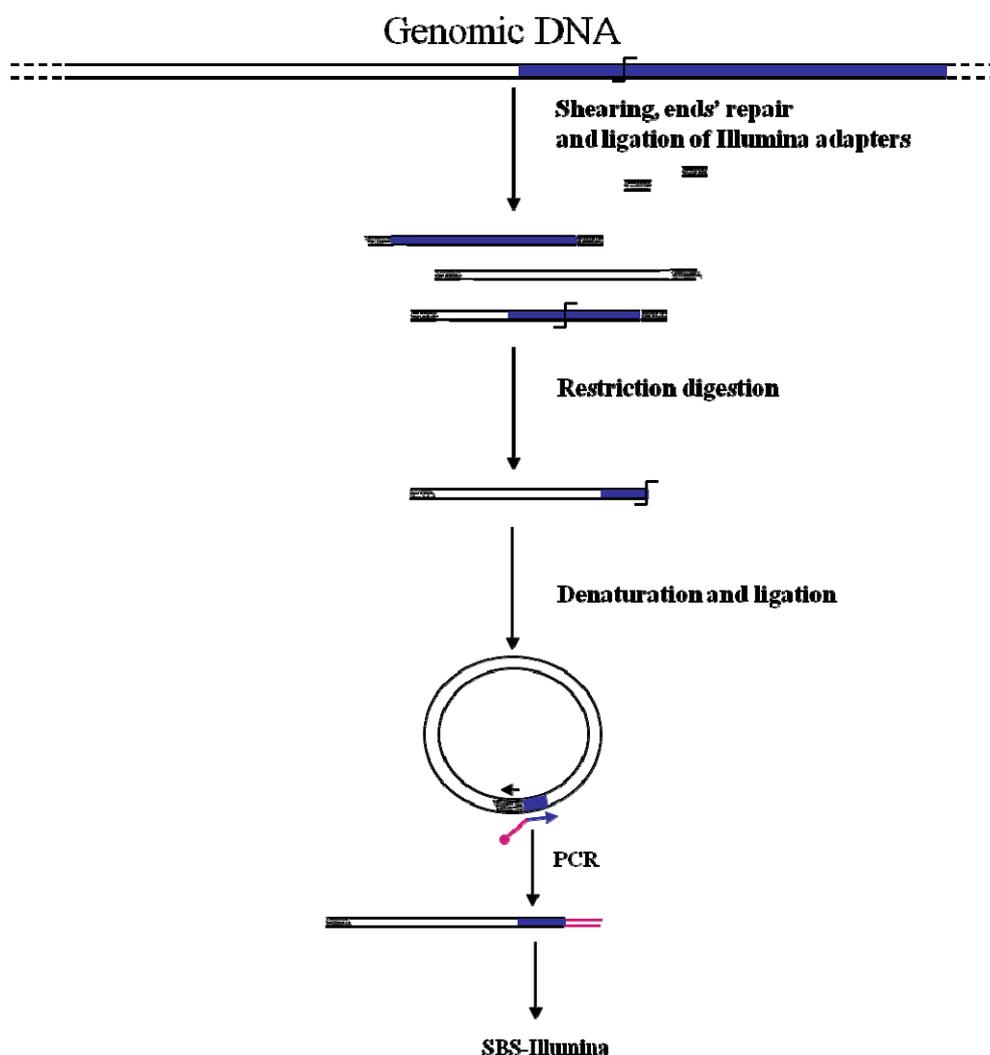
Brett *et al.* [26] designed an Illumina-GW procedure for the analysis of *Sleeping Beauty (SB)* induced tumors in mice. The method is based on the ligation of adapters to genomic DNA fragments obtained by digestion with restriction enzymes (*AluI* or *NlaIII*). After a first PCR with transposon and adapter specific primers, a nested PCR is performed with primers provided of barcodes and tags for sequencing on Illumina platforms. Regardless of the NGS technology, the approach is similar to that described by Wang *et al.* [13] and illustrated in Figure 1B.

**Figure 4.** Main steps of *MmeI*-assisted cassette-PCR GW and SBS-Illumina association strategy. Blue and white bars correspond to known and unknown sequences, respectively. An open vertical arrow indicates the *MmeI* recognition site. The  $\lrcorner$  symbol indicates the *MmeI* digestion site. A black horizontal arrow indicates adapter specific primers; a blue arrow represents primers designed in the known region. Barcodes and Illumina sequencing traits of longer amplification primers are indicated in the figure.



An innovative NGS-GW approach is that reported by Gallagher *et al.* [27] to track large numbers of transposon mutants of *Pseudomonas aeruginosa*. In this case authors associated an inverted-PCR GW approach (firstly described by Triglia *et al.* [28]) to the Illumina NGS technology (Figure 5). Genomic DNA from a mutant pool is sheared, end-repaired and ligated to one Illumina adaptor. DNA fragments are then digested by a proper restriction enzyme cutting the transposon near the junction with the unknown DNA. After size-selection, restriction fragments are denatured and single-strand DNA self-ligated. A second PCR step with divergent primers corresponding to the adaptor and to the transposon allows the identification of transposon/genome junctions. Illumina sequencing can be performed by introducing into the transposon primer the required Illumina adaptor sequence. The procedure was tested by screening a pool of 100,000 mutants for the identification of tobramycin resistance genes. A total of 117 resistance genes were identified, including previously unknown ones.

**Figure 5.** Main steps of Inverted-PCR GW and SBS-Illumina association strategy. Blue and white bars correspond to known and unknown sequences, respectively. An open vertical arrow indicates the *MmeI* recognition site. The  $\lrcorner$  symbol indicates a restriction site. A black horizontal arrow indicates adapter specific primers; a blue arrow represents primers designed in the known region. The Illumina sequencing trait of longer amplification primers is indicated in the figure.



#### 4. Other Possible NGS-GW Pairing and General Conclusions

The possibility to combine other NGS and GW strategies, in addition to those already reported, is examined in this paragraph. This analysis, besides constituting a preliminary base for future applications, also gives the opportunity to draw some general conclusions about the NGS and GW strategies to combine for optimal results.

Taking into account that all the GW methods have a final PCR amplification step, NGS methods relying on the synthesis of large arrays of amplicons as substrate for sequencing can in principle be combined. Even if currently there are no articles reporting the application of the SOLiD and Ion Torrent NGS technologies to GW methods, the possibility to construct large barcoded libraries has been recently illustrated for both methods, making them eligible for large, multiplex, GW applications. For example, the construction of barcoded libraries for SOLiD sequencer was reported by Farias-Hesson *et al.* [29], who prepared 32 libraries by ligation of end-repaired sheared DNA molecules to SOLiD adaptors, one of which provided with 6-bp barcode. This procedure should be also adaptable to multiple sequence analysis of the same targeted region (*i.e.*, a transposon) amongst several biological samples (see also the Applied Biosystem Application note “SOLiD System Barcoding” [30]). Additional examples of the use of barcoding to analyze the same targeted region in several biological samples by SOLiD technology can be found in some recent papers [31–34]. The possibility to apply the Ion Torrent sequencing technology to GW can be deduced from the genotyping-by-sequencing approach exemplified for the complex genomes of barley and wheat and available at the Invitrogen website [35]. The lower cost of the Personal Genome Machines by Ion Torrent compared to SOLiD platforms, together with their reduced times for amplicon sequencing, can make them probably preferred for most NGS applications, including GW.

From the analysis of the GW-NGS studies reported in the previous paragraphs, it comes out that pyrosequencing has been exclusively used for eukaryotic genomes, while the SBS-Illumina approach has been adopted for both prokaryotic and eukaryotic genomes. However, a definitive indication about the most appropriate NGS technology to use in GW insertional mutation analysis is not really relevant at this time, since all the NGS approaches reported above have larger sequencing capacity than that needed for identification of insertion sites. This is true even for high-copy transposons in eukaryotic genomes, as the *Mu* transposon in maize, which has been tracked by both GW-pyrosequencing [14,15] and GW-Illumina approaches [22]. Accordingly, in all the cases the introduction of DNA barcodes allowed the analysis of multiple samples in single sequencing experiments. In the case of *SB*-induced tumors, however, Brett *et al.* [26] reported about the higher capacity of insertion sites detection by SBS-Illumina compared to 454-pyrosequencing. Even if some experimental details were different, comparison of the two sequencing strategies in the analysis of ligation-mediated PCR libraries showed an increased sensitivity of the Illumina approach respect to pyrosequencing of about 50 fold, allowing the detection of rare transposon insertion events. In the case of bacteria with high GC-content, Ion Torrent approach has been found more reliable than Illumina, with a more stable quality of sequencing data (reviewed in Liu *et al.* [36]).

Also for the GW method to combine with NGS technologies, data available indicate that both restriction-based and extension-based approaches are suitable for adaptation to NGS technologies. The first case is exemplified by the cassette-ligation (Figure 1), its “shearing-adapted” version (Figure 3),

and by the self-ligation (Figure 5) methods. The latter approach is shown in Figure 2. Association of primer-based GW methods to NGS approaches appears more difficult. Primer-based GW strategies start with a preliminary PCR carried out with a primer containing degenerate sequences (possibly annealing in the unknown region of the genome) and a primer corresponding to the known sequence (*i.e.*, a gene region or an insertion sequence). Subsequently PCR products are further selected by hemi-nested PCR amplifications using a second primer specific for the known region. Even if in this last step nested primer containing DNA barcodes and NGS sequencing regions can be designed, making the association to NGS approaches theoretically possible, currently there are no reports about the combined use of the two methods.

Another critical parameter to be considered in the choice of the GW approach is its sensitivity. Some specific comparative studies have been carried out and discussed in a recent review article [1]. Generally speaking, however, the use of solid-phase purification of biotinylated DNA fragments has highly contributed to increasing the sensitivity of GW methods. For example in studies on the integration of foreign DNA into salmon genome, it has been found that classical anchored PCR amplifications, combined with solid-phase purification, allow identification of about two copies of the target sequence in 25 ng of DNA [37]. Several of the above reported GW-NGS approaches include streptavidin purification of biotinylated DNA molecules [9,10,12,22].

Finally, the need of proper bioinformatic resources for the analysis of GW-NGS data cannot be disregarded. A typical analysis pipeline includes: sorting of sequences according to the incorporated barcode for the identification of sequences belonging to individual amplicons; removal of foreign genome sequences (linkers, transposons, vectors); clustering and counting of identical sequences and mapping of insertion sites. Good examples of bioinformatic analysis with description of the employed tools and filtering parameters are given in several articles on GW-NGS applications [7,9,12,14,15,20–23,26,27].

In conclusion, GW-NGS approaches have the potentiality to make insertional mutagenesis a high throughput screening approach of wide use in functional genomics, replacing more classical methods such as microarray hybridization. In addition, the combined methods showed interesting applications in medicine for the study of the integration mechanisms of retrovirus and DNA vector for gene therapy.

## Acknowledgments

Work partly supported by the Italian Ministry of Instruction University and Research: PRIN 20087ATS57 “Food allergens”. C. Leoni was recipient of a fellowship from the Italian Consortium for Biotechnologies. Authors wish to thank Graziano Pesole (Department of Biosciences, Biotechnology and Pharmacological Sciences, University of Bari) for his helpful comments.

## References

1. Leoni, C.; Volpicella, M.; de Leo, F.; Gallerani, R.; Ceci, L.R. Genome walking in eukaryotes. *FEBS J.* **2011**, *278*, 3953–3977.
2. Tonooka, Y.; Fujishima, M. Comparison and critical evaluation of pcr-mediated methods to walk along the sequence of genomic DNA. *Appl. Microbiol. Biotechnol.* **2009**, *85*, 37–43.

3. Kotik, M. Novel genes retrieved from environmental DNA by polymerase chain reaction: Current genome-walking techniques for future metagenome applications. *J. Biotechnol.* **2009**, *144*, 75–82.
4. Shendure, J.; Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145.
5. Metzker, M.L. Sequencing technologies—The next generation. *Nat. Rev. Genet.* **2010**, *11*, 31–46.
6. Wang, G.P.; Ciuffi, A.; Leipzig, J.; Berry, C.C.; Bushman, F.D. HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **2007**, *17*, 1186–1194.
7. Gawronski, J.D.; Wong, S.M.; Giannoukos, G.; Ward, D.V.; Akerley, B.J. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for haemophilus genes required in the lung. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 16422–16427.
8. Van Opijnen, T.; Camilli, A. Genome-wide fitness and genetic interactions determined by tn-seq, a high-throughput massively parallel sequencing method for microorganisms. *Curr. Protoc. Microbiol.* **2010**, doi:10.1002/9780471729259.mc01e03s19.
9. Paruzynski, A.; Arens, A.; Gabriel, R.; Bartholomae, C.C.; Scholz, S.; Wang, W.; Wolf, S.; Glimm, H.; Schmidt, M.; von Kalle, C. Genome-wide high-throughput integrome analyses by nrlam-pcr and next-generation sequencing. *Nat. Protoc.* **2010**, *5*, 1379–1395.
10. Bartholomae, C.C.; Glimm, H.; von Kalle, C.; Schmidt, M. Insertion site pattern: Global approach by linear amplification-mediated PCR and mass sequencing. *Methods Mol. Biol.* **2012**, *859*, 255–265.
11. Rothberg, J.M.; Hinz, W.; Rearick, T.M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J.H.; Johnson, K.; Milgrew, M.J.; Edwards, M.; *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **2011**, *475*, 348–352.
12. Vandenbussche, M.; Janssen, A.; Zethof, J.; van Orsouw, N.; Peters, J.; van Eijk, M.J.; Rijpkema, A.S.; Schneiders, H.; Santhanam, P.; de Been, M.; *et al.* Generation of a 3D indexed petunia insertion database for reverse genetics. *Plant J.* **2008**, *54*, 1105–1114.
13. Wang, G.P.; Garrigue, A.; Ciuffi, A.; Ronen, K.; Leipzig, J.; Berry, C.; Lagresle-Peyrou, C.; Benjelloun, F.; Hacein-Bey-Abina, S.; Fischer, A.; *et al.* DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. *Nucleic Acids Res.* **2008**, *36*, e49.
14. Liu, S.; Dietrich, C.R.; Schnable, P.S. Dla-based strategies for cloning insertion mutants: Cloning the gl4 locus of maize using mu transposon tagged alleles. *Genetics* **2009**, *183*, 1215–1225.
15. Liu, S.; Yeh, C.T.; Ji, T.; Ying, K.; Wu, H.; Tang, H.M.; Fu, Y.; Nettleton, D.; Schnable, P.S. Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet.* **2009**, *5*, e1000733.
16. Guo, Y.; Levin, H.L. High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in *Schizosaccharomyces pombe*. *Genome Res.* **2010**, *20*, 239–248.
17. Mueller, P.R.; Wold, B. *In vivo* footprinting of a muscle specific enhancer by ligation mediated pcr. *Science* **1989**, *246*, 780–786.
18. Pfeifer, G.P.; Steigerwald, S.D.; Mueller, P.R.; Wold, B.; Riggs, A.D. Genomic sequencing and methylation analysis by ligation mediated pcr. *Science* **1989**, *246*, 810–813.

19. Schmidt, M.; Zickler, P.; Hoffmann, G.; Haas, S.; Wissler, M.; Muessig, A.; Tisdale, J.F.; Kuramoto, K.; Andrews, R.G.; Wu, T.; *et al.* Polyclonal long-term repopulating stem cell clones in a primate model. *Blood* **2002**, *100*, 2737–2743.
20. Langridge, G.C.; Phan, M.D.; Turner, D.J.; Perkins, T.T.; Parts, L.; Haase, J.; Charles, I.; Maskell, D.J.; Peters, S.E.; Dougan, G.; *et al.* Simultaneous assay of every salmonella typhi gene using one million transposon mutants. *Genome Res.* **2009**, *19*, 2308–2316.
21. Li, J.; Zhang, J.M.; Li, X.; Suo, F.; Zhang, M.J.; Hou, W.; Han, J.; Du, L.L. A piggybac transposon-based mutagenesis system for the fission yeast *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **2011**, *39*, e40.
22. Williams-Carrier, R.; Stiffler, N.; Belcher, S.; Kroeger, T.; Stern, D.B.; Monde, R.A.; Coalter, R.; Barkan, A. Use of illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy mutator lines of maize. *Plant J.* **2010**, *63*, 167–177.
23. Urbanski, D.F.; Malolepszy, A.; Stougaard, J.; Andersen, S.U. Genome-wide *lorel1* retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. *Plant J.* **2012**, *69*, 731–741.
24. Van Opijnen, T.; Bodi, K.L.; Camilli, A. Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* **2009**, *6*, 767–772.
25. Goodman, A.L.; McNulty, N.P.; Zhao, Y.; Leip, D.; Mitra, R.D.; Lozupone, C.A.; Knight, R.; Gordon, J.I. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **2009**, *6*, 279–289.
26. Brett, B.T.; Berquam-Vrieze, K.E.; Nannapaneni, K.; Huang, J.; Scheetz, T.E.; Dupuy, A.J. Novel molecular and computational methods improve the accuracy of insertion site analysis in sleeping beauty-induced tumors. *PLoS One* **2011**, *6*, e24668.
27. Gallagher, L.A.; Shendure, J.; Manoil, C. Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using tn-seq. *MBio* **2011**, *2*, e00315-10.
28. Triglia, T.; Peterson, M.G.; Kemp, D.J. A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res.* **1988**, *16*, 8186.
29. Farias-Hesson, E.; Erikson, J.; Atkins, A.; Shen, P.; Davis, R.W.; Scharfe, C.; Pourmand, N. Semi-automated library preparation for high-throughput DNA sequencing platforms. *J. Biomed. Biotechnol.* **2010**, *2010*, 617469.
30. SOLiD™ System Barcoding. Available online: [http://www3.appliedbiosystems.com/cms/groups/mcb\\_marketing/documents/generaldocuments/cms\\_057554.pdf](http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_057554.pdf) (accessed on 1 August 2012).
31. Harakalova, M.; Mokry, M.; Hrdlickova, B.; Renkens, I.; Duran, K.; van Roekel, H.; Lansu, N.; van Roosmalen, M.; de Bruijn, E.; Nijman, I.J.; *et al.* Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. *Nat. Protoc.* **2011**, *6*, 1870–1886.
32. Lotta, L.A.; Wang, M.; Yu, J.; Martinelli, I.; Yu, F.; Passamonti, S.M.; Consonni, D.; Pappalardo, E.; Menegatti, M.; Scherer, S.E.; *et al.* Identification of genetic risk variants for deep vein thrombosis by multiplexed next-generation sequencing of 186 hemostatic/pro-inflammatory genes. *BMC Med. Genomics* **2012**, *5*, 7.

33. Harakalova, M.; Nijman, I.J.; Medic, J.; Mokry, M.; Renkens, I.; Blankensteijn, J.D.; Kloosterman, W.; Baas, A.F.; Cuppen, E. Genomic DNA pooling strategy for next-generation sequencing-based rare variant discovery in abdominal aortic aneurysm regions of interest-challenges and limitations. *J. Cardiovasc. Transl. Res.* **2011**, *4*, 271–280.
34. Tu, J.; Ge, Q.; Wang, S.; Wang, L.; Sun, B.; Yang, Q.; Bai, Y.; Lu, Z. Pair-barcode high-throughput sequencing for large-scale multiplexed sample analysis. *BMC Genomics* **2012**, *13*, 43.
35. Genotyping by Sequencing (GBS) on Barley Using Ion PGM™ Sequencer: A Feasibility Study. Available online: <http://www.invitrogen.com/etc/medialib/images/agricultural-biotechnology/pdf.Par.0661.File.tmp/Genotyping-by-sequencing-on-barley-Ion-PGM.pdf>. (accessed on 1 August 2012).
36. Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L.; Law, M. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, *2012*, 251364.
37. Nielsen, C.R.; Berdal, K.G.; Holst-Jensen, A. Anchored pcr for possible detection and characterisation of foreign integrated DNA at near single molecule level. *Eur. Food Res. Technol.* **2008**, *226*, 949–956.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).