Oral Oncology 50 (2014) 247–253

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Review

Next generation sequencing and its application in deciphering head and neck cancer

Maryam Jessri [a,b], Camile S Farah [a,b,*]

[a] The University of Queensland, UQ Centre for Clinical Research, Herston, Qld 4029, Australia
[b] The University of Queensland, School of Dentistry, Brisbane, Qld 4000, Australia

ARTICLE INFO

Article history:
Received 14 October 2013
Received in revised form 13 December 2013
Accepted 14 December 2013
Available online 15 January 2014

Keywords:
Head and neck cancer
Oral cancer
Next generation sequencing
RNA-seq
Personalized cancer therapy

SUMMARY

Head and neck squamous cell carcinoma (HNSCC) are a group of heterogeneous tumours mainly attributable to tobacco use, alcohol consumption and infection with human papillomavirus. Based on the stage of cancer at the time of diagnosis, patients are managed by surgery, radiotherapy, chemotherapy or a combination of these. Early diagnosis usually improves patient prognosis.

Since their first commercial application in 2005, next generation sequencing (NGS) platforms are rapidly changing the face of basic science laboratories; however prior to progressing to clinical applications, clinicians should carefully examine currently available data and guidelines for technical and ethical matters concerning NGS. In this review, we compare various commercially available NGS platforms, with special consideration given to their clinical application in the management of HNSCC.

Introduction

Early diagnosis and efficient management improves the quality of life of patients with head and neck cancer (HNC) and limits their impairment. Despite advances in imaging and diagnostic techniques and increased use of chemotherapy and radiotherapy, the 5 year survival rate for head and neck squamous cell carcinoma (HNSCC) has remained largely unchanged for the past three decades [1,2]. The tumour is often locally advanced (stage IV) with significant lymph node involvement at the time of diagnosis [3].

An environmentally caused cancer, with tobacco use and alcohol consumption being the main causative factors [4], HNSCC has been recently linked to infection with human papillomavirus (HPV). HPV-induced HNSCCs are generally limited to the oropharynx and are most frequently seen in non-smoker, non-drinker individuals [5–7]. Despite positive expression of p16 and contrary to squamous cell carcinomas (SCC) of the oropharynx, squamous cell carcinomas of the tongue rarely harbour HPV [8–13]. SCC of the tongue comprises a major subgroup of oral squamous cell carcinoma (OSCC) and recently it has been noted that a growing proportion of these cases occur in young patients lacking traditional HNSCC risk factors [14]. OSCC in young low-risk patients is more aggressive and has a distinctive clinical and histopathological pattern which requires extensive care and management [10,15,16]. Although more than 90% of HNC are squamous cell carcinoma, recent studies have revealed this tumour to have a very heterogeneous nature [17,18]. These findings are highly suggestive of a role for genetic factors in carcinogenesis of HNSCC and were confirmed by DNA and RNA profiling techniques demonstrating that different subclasses of SCC have different prognoses and recurrence rates [19–21].

Next generation sequencing (NGS) accelerates the process of studying DNA and RNA by generating digital and quantifiable data that can be mapped back to the genome or transcriptome. By comparing the tumour sequence to germline, NGS can potentially reveal somatic-genetic alterations in cancer genomes.

Considering the heterogeneity of HNSCC, NGS could be the perfect means by which to study and understand the genomic alterations that result in formation of this group of tumours. Heterogeneity of HNSCC, in addition to the increasing number of young patients with OSCC, and finally involvement of HPV as a causative factor make the enormous gap in our knowledge in regard to HNSCC apparent. Findings from NGS studies of HNSCC will help us better understand the genetic aspects of a tumour traditionally considered environmental. Although high cost and complicated analysis limit the routine application of NGS in the current management of HNSCC, commercial NGS platforms are becoming increasingly more cost effective and user-friendly; making their clinical use more feasible. In this review we examine different NGS platforms and their implications in oral oncology with a view to their limitations and future applications in a clinical setting.

* Corresponding author. Address: UQ Centre for Clinical Research, Royal Brisbane & Women's Hospital, The University of Queensland, Herston, Qld 4029, Australia. Tel.: +61 7 3346 6030; fax: +61 7 3346 6098.
E-mail address: c.farah@uq.edu.au (C.S Farah).

1368-8375/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.oraloncology.2013.12.017

Next generation sequencing platforms

The first DNA sequencing method was introduced by Sanger et al., and was based on incorporation of slightly modified nucleotides for chain elongation by DNA polymerase [22]. The Sanger method was laborious, time consuming, and error prone given the number of manual laboratory-based steps required and hand entry of resulting sequences. Beginning in 2005, new sequencing methods, commonly referred to as "next generation sequencing" have revolutionized traditional Sanger sequencing. Unlike Sanger sequencing, NGS platforms do not require a cloning step; instead they use synthetic DNA fragments (adapters), specifically designed for each platform, to amplify the DNA library on a solid support matrix followed by cyclic sequencing. Furthermore, while different steps of sequencing and signal detection are consecutive in the Sanger method, commercial NGS instruments are capable of performing these simultaneously; accordingly they are referred to as parallel sequencing. Finally, the read length in Sanger sequencing is limited by gel electrophoresis-related issues while in NGS it is determined by the signal-to-noise ratio [23]. With underlying differences in principles and chemistry, the Sanger method and NGS are prone to different types of error and as a result, despite higher accuracy, data from NGS platforms are still verified by the Sanger method [24].

The terms second-generation and third-generation sequencing have been used to describe the shift from dideoxy Sanger technique to NGS. However the evolution from PCR-based sequencing methods to a cycle-free chemistry which sequences single molecules is characteristic of progression to third-generation sequencing methods [25]. Although a detailed description of different technologies is beyond the scope of this review, Table 1 presents a concise summary of current commercially available NGS instruments.

Second-generation sequencing platforms have similar library preparation methods with the final aim of ligating platform-specific adapters to the terminal ends of DNA fragments. They share the processes of clonal amplification of DNA molecules spacially separated on a solid matrix, and cyclic addition of sequencing chemistries. Despite this, NGS platforms differ in the methods implemented to achieve these processes.

Roche 454

The first commercial massively parallel DNA sequencing platform was 454 pyrosequencing (Qiagen, Venlo, Netherlands). All Roche platforms use emulsion PCR which is based on adding equimolar concentrations of beads and template library to an emulsion of water and oil such that each emulsion droplet potentially contains a single bead and a single DNA molecule. DNA libraries that are attached to primer-coated agarose beads via 454 specific adaptors are amplified to coat each bead with one million copies of that single DNA fragment. Enriched beads are fixed in a picotiter plate and pure nucleotide solutions are added in a step-wise fashion. Incorporation of each nucleotide results in pyrophosphate release which initiates a series of reactions ultimately producing light. The chemiluminescent signal indicates base incorporation and the intensity of light is proportional to the number of nucleotides added. The signal is located and visualized by a fiberoptic-coupled imaging camera [26]. Roche 454FLX+ distinguishes itself by its significantly longer reads (700–1000 bp as of October 2013). Despite the longer reads and being relatively fast, application of this system is limited by high error rate in genetic regions rich in homopolymer repeats greater than seven bases in length [27]. In addition, this platform has a significantly lower output compared to the other second-generation sequencing platforms.

Table 1
Current commercially available NGS instruments.

Platforms	Sequencing principle	Read length (base pair)	Run Time	Max output	Pros	Cons	Application
Roche 454 FLX+	Pyrosequencing	700	23 h	0.7 GB	High throughput, long reads, short run time, high coverage	High error rate in homopolymers regions, long hand-on time	De novo whole genome sequencing of microbes, exome sequencing
Roche 454 GS Junior (BT)	Pyrosequencing	400	10 h	0.035 GB	Long reads, short run time	High error rate in homopolymers regions, long hand-on time	De novo whole genome sequencing of microbes, exome sequencing
Illumina® HiSeq™ 2500/1500	Sequencing by synthesis	36/50/100	11 days 27 h (rapid run)	600 GB	High throughput, cost-effectiveness	Short read length; long run time; high error rate	Human whole genome sequencing, exome sequencing, RNA-seq, methylation
Illumina® MiSeq™ (BT)	Sequencing by synthesis	25/36/100	4–27 h	8.5 GB	High throughput, cost-effectiveness, short run time, high coverage	Short read length	Microbial discovery, exome sequencing, targeted capture
Life Technologies™ SOLiD™ 5500XL	Sequencing by ligation	75+35	7 days to 4 weeks	180 GB	Very high throughput, low error rate	Long run times, short reads, complexity of library preparation	Human whole genome sequencing, whole exome sequencing, RNa-seq, methylation
Life Technologies™ Ion PGM™(BT)	Hydrogen ion sensitive transistor	35/200/400	2 h	2 GB	Short run time, low cost per sample	High reagent cost, high error rate in homopolymers region	Exome sequening, targeted gene sequencing, microbial discovery
Life Technologies™ Ion Proton™ Chipl/ II(BT)	Hydrogen ion sensitive transistor	Up to 200	2 h	10 GB/ 100 GB	Short run times, flexible chip reagents	High reagent cost, high error rate in homopolymers region, limited body of literature to support	Exome sequening, targeted gene sequencing, microbial discovery
Pacific Biosciences Inc. PacBio RS II (TGS)	Single molecule real-time sequencing	250–10,000	30–120 min	217 Mb	Short run times, no PCR amplification, long reads, low cost	Not optimized for human genome,	Virus, bacteria, lower eukaryotes genome sequencing, target sequencing

BT: Bench-top instrument.
TGS: Third generation sequencing.