Cancer Genetics 209 (2016) 57–69

Cancer Genetics

Overview of recurrent chromosomal losses in retinoblastoma detected by low coverage next generation sequencing

A.J. García-Chequer [a], A. Méndez-Tenorio [b], G. Olguín-Ruiz [b], C. Sánchez-Vallejo [b], P. Isa [c], C.F. Arias [c], J. Torres [a], A. Hernández-Angeles [a], M.A. Ramírez-Ortiz [d], C. Lara [d], M.L. Cabrera-Muñoz [d], S. Sadowinski-Pine [d], J.C. Bravo-Ortiz [e], G. Ramón-García [e], J. Diegopérez-Ramírez [e], G. Ramírez-Reyes [e], R. Casarrubias-Islas [e], J. Ramírez [f], M.A. Orjuela [g], M.V. Ponce-Castañeda [a,*]

[a] *Unidad de Investigación Médica en Enfermedades Infecciosas, Centro Médico Nacional SXXI, Instituto Mexicano del Seguro Social, México D.F., Mexico;* [b] *Lab. Bioinformática Genómica, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, México D.F., Mexico;* [c] *Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico;* [d] *Hospital Infantil de México Federico Gómez, México D.F., Mexico;* [e] *Hospital de Pediatría, CMN SXXI, Instituto Mexicano del Seguro Social, México D.F., Mexico;* [f] *Unidad de Microarreglos, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, México D.F., Mexico;* [g] *Columbia University, New York, USA*

Genes are frequently lost or gained in malignant tumors and the analysis of these changes can be informative about the underlying tumor biology. Retinoblastoma is a pediatric intraocular malignancy, and since deletions in chromosome 13 have been described in this tumor, we performed genome wide sequencing with the Illumina platform to test whether recurrent losses could be detected in low coverage data from DNA pools of Rb cases. An *in silico* reference profile for each pool was created from the human genome sequence GRCh37p5; a chromosome integrity score and a graphics 40 Kb window analysis approach, allowed us to identify with high resolution previously reported non random recurrent losses in all chromosomes of these tumors. We also found a pattern of gains and losses associated to clear and dark cytogenetic bands respectively. We further analyze a pool of medulloblastoma and found a more stable genomic profile and previously reported losses in this tumor. This approach facilitates identification of recurrent deletions from many patients that may be biological relevant for tumor development.

Keywords Pediatric tumors, retinoblastoma, medulloblastoma, next-generation sequencing, low coverage, losses

Introduction

Retinoblastoma (Rb) is an intraocular malignancy that develops in early childhood and originates in the retina. About 60–70% of all cases are unilateral and 30–40% bilateral and unfortunately surgical removal of the eye is often the first line of treatment (1). Retinoblastoma is considered a robust clinical model for genetic predisposition to development of cancer, because offspring of patients with bilateral Rb are also affected by the disease exhibiting a dominant autosomic pattern of inheritance. In addition, patients with bilateral Rb also have a significantly higher lifetime risk to develop secondary malignancies. Early cytogenetic analyses of this tumor demonstrated recurrent losses of the long arm in chromosome 13 (2–6), and these gross deletions allowed the localization and cloning of the retinoblastoma susceptibility gene RB1, the first tumor suppressor gene discovered (7,8).

Received June 25, 2015; received in revised form September 1, 2015; accepted December 3, 2015.

* Corresponding author.
E-mail address: vponce@ifc.unam.mx.

2210-7762/$ - see front matter © 2015 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cancergen.2015.12.001

Genetic alterations are hallmarks of malignant processes. In particular recurrent genetic alterations can be detected across different patients, and such recurrence is an indication that these are driving events for tumor development (9). Recurrent losses may indicate the presence of tumor suppressor genes within the lost regions, as was the case for RB1, while gained or amplified loci may translate in increased expression of amplified genes, often functioning as oncogenes like ERBB2 (also known as Her-2/neu) gene in breast cancer (10,11).

Our ability to detect oncogenic driver genes whether lost or gained has been limited by the resolution of available methods to detect these key recurrent gains or losses. In retinoblastoma recurrent chromosome gains and losses other than in chromosome 13, have been described although with variable frequencies using karyotype, FISH (Fluorescence In Situ Hybridization), CGH (Comparative Genome Hybridization), SNPs (Single Nucleotide Polymorphism), qPCR and whole genome sequence (12–19), including gains at 1q, 6p, 17q and 19q and losses at 16q and X, 2p, 5q. Thus, more oncogenes and tumor suppressor genes are thought to be or have been located in those altered regions (20–22).

High density data technologies like microarrays and massive parallel sequencing are new powerful tools able to detect many of these variations in gene copy number (23–25). Next generation sequencing (NGS) is a high throughput technology used with increasing frequency given to decreasing costs. NGS allows the sequencing of complete genomes through the making and sequencing of DNA libraries that are overlapping discrete fragments of the genome. The sequences obtained are a large collection of so called “reads”. According with the length and number of reads obtained per sample, it is nearly possible to reconstruct complete human genomes. Depth and coverage depend on the number of reads obtained in relation to the genome length, and it refers to the number of times a given region has been covered by different reads. Depending on the depth of coverage of massive sequencing data, different types of analyses requiring different resolution can be performed with NGS data, including genome reconstruction, determination of mutations, polymorphisms, splicing events, isoforms analysis, gains and losses and translocations (26–28).

Because this technology is relatively new, methods for analysis, bioinformatic algorithms and validation to detect gains and losses are still in development (28–30), and it has not yet been thoroughly established what are the limits on the depth of sequencing data that can be useful for the identification of recurrent gains and losses (31).

Since retinoblastoma is a tumor that has been widely studied in terms of recurrent gains and losses, we chose to utilize 8 Rb cases as biological replicas to validate a procedure using low coverage-massive sequencing data to detect recurrent losses. To do this we used for sequencing pooled DNA from different retinoblastoma cases, assuming that detection of losses in a pool indicates recurrence, since all the tumors in the pool must contribute to the loss in order to be detected. In contrast gains detected cannot be called recurrent with this approach because it is impossible to discriminate the individual contribution for gains in such a pool. We also developed a bioinformatic strategy to replace the complexities and costs of sequencing a normal tissue sample with an *in silico* reference for each tumor pool sequenced. This allowed us to use $\log_2$ ratios of the number of reads between tumor and reference to detect and quantify gains and losses. We further tested this approach analyzing the profile of pooled samples of medulloblastoma, another malignant primitive neuroectodermal tumor that is histomorphologically similar to retinoblastoma. In both tumors we found losses and gains previously reported, in retinoblastoma many more gross recurrent and non random losses were identified in most chromosomes, while in medulloblastoma we found very few gross alterations. Interestingly in retinoblastoma we found an association between gross deletions and dark cytogenetic bands and gross amplifications and the corresponding clear cytogenetic bands.

Materials and methods

DNA samples

DNA from eight retinoblastoma and five medulloblastoma tumor samples was used for the next generation sequencing analyses. Two DNA pools of four Rb samples each, as technical replicas, and one DNA pool containing five Mb patients, were sequenced. Each Rb pool was gender balanced containing 2 boys and 2 girls and the Mb pool consisted of 4 boys and 1 girl (see Table 1 for further details).

All patients were newly diagnosed and tumor samples were collected at time of surgery, prior to any adjuvant therapy, from children diagnosed and treated at Hospital de Pediatría from Instituto Mexicano del Seguro Social (IMSS) and Hospital Infantil de México Federico Gómez from Secretaría de Salud in Mexico City. Retinoblastoma samples were collected as part of a separate study which included demographic and clinical data collection as previously described (32). Tumor tissues were collected under informed written consent from their parents and as part of studies approved by the Scientific and Ethics Review Boards from each participating institution. Patient characteristics are shown in Table 1. DNA was extracted from the fresh frozen tumor tissues by a non-organic method using Qiagen and 5PRIME (DNeasy Blood and Tissue Kit,

Table 1 Clinical and demographic features of participating retinoblastoma and medulloblastoma patients included in the DNA pools

Pool	ID	Gender	Laterality	St-Jude stage
Rb 1	F86	Female	Unilateral	III-c
Rb 1	F78	Male	Unilateral	II-c
Rb 1	F37	Female	Bilateral	II-d
Rb 1	F73	Male	Bilateral	II-c
Rb 2	F54	Male	Unilateral	II-c
Rb 2	F-1	Female	Bilateral	II-c
Rb 2	F81	Female	Unilateral	II-a
Rb 2	F168	Male	Bilateral	II-a

Pool	ID	Gender	Histology
Mb	FM 1	Male	Classic IV (WHC)
Mb	FM 15	Female	Classic IV (WHC)
Mb	FM 3	Male	Classic IV (WHC)
Mb	FM 17	Male	Desmoplastic
Mb	FM 22	Male	Classic IV (WHC)