Mutational Landscape of Basal Cell Carcinomas by Whole-Exome Sequencing

Shyam S. Jayaraman[1], David J. Rayhan[2], Salar Hazany[2] and Michael S. Kolodney[1]

Recent advances in sequencing technology allow genome-scale approaches to cancer mutation discovery. Such data-intensive methods have been applied to cutaneous squamous cell carcinomas (SCCs) and melanomas but have not, to our knowledge, been applied to basal cell carcinomas (BCCs). We used whole-exome sequencing to characterize the mutational landscape of sporadic BCCs. We show that BCCs are the most mutated type of human cancer. Tumors from anatomical regions with chronic UV exposure were associated with higher mutation rates than those with intermittent exposure. The majority of all mutations (75.7%) were UV signature. Using a conventional binomial probability model, several genes were found mutated significantly. However, this model assumes a uniform distribution of mutations throughout the genome. We also used a more stringent approach called InVEx that uses a permutation-based framework to pick drivers from passengers. After correction for multiple hypothesis testing, InVEx identified only *PTCH1* (*Patched 1*) as having a significant functional mutation burden. We also found three genes, *STAT5B*, *CRNKL1*, and *NEBL*, with mutational hot spots at a single base in 3 of 12 tumors sequenced. Our findings support the central role of *PTCH1* mutations in BCCgenesis. Moreover, our discovery of the uniquely high number of mutations in this tumor may lend insight into its biological behavior.

Journal of Investigative Dermatology (2014) **134,** 213–220; doi:10.1038/jid.2013.276; published online 25 July 2013

INTRODUCTION

Basal cell carcinoma (BCC) is the most common human cancer, accounting for 80–90% of all primary skin cancers (Scotto *et al.*, 1983; Iwasaki *et al.*, 2012). UV exposure is the most important risk factor in basal cell carcinogenesis. However, crucial insight into the biology of BCC came from linkage analysis of families with a rare inherited form of the disease.

Studies on Gorlin's syndrome, an autosomal dominant disorder predisposing to multiple BCCs, initially demonstrated a pivotal role for the sonic hedgehog (SHH) pathway in BCC. Pioneering linkage analysis of families with this syndrome localized a putative tumor-suppressor gene to 9q22.3–q31 (Farndon *et al.*, 1992; Gailani *et al.*, 1992; Reis *et al.*, 1992). Further studies (Chenevix-Trench *et al.*, 1993; Shanley *et al.*, 1995) led to the identification of mutations in the human homolog of the Drosophila patched protein as the cause of Gorlin's syndrome (Hahn *et al.*, 1996; Johnson *et al.*, 1996). Further research has confirmed the central role of the SHH/PTCH1/SMO pathway in sporadic BCCs. PTCH1 (Patched 1) normally acts to keep the activity of SMO (Smoothened) in check, by playing an inhibitory role. The binding of SHH to the extracellular domain of PTCH1 relieves SMO inhibition, leading to activation of the GLI transcription factors that regulate the transcription of hedgehog target genes. Recently, the Patched pathway became a therapeutic target with the introduction of a clinically effective small-molecule antagonist of the SMO receptor—Vismodegib (Rudin, 2012).

Numerous studies of sporadic BCCs have reported point mutations, copy-loss loss of heterozygosity (LoH), and copy-neutral LoH (due to uniparental disomy) in *PTCH1* (Teh *et al.*, 2005; Santos *et al.*, 2011). For example, one study (Teh *et al.*, 2005) reported a highly conserved LoH region at 9q21–q31 found in 13 of 14 (93%) BCCs with uniparental disomy being found in 5 of 13 (38%) BCCs. The *de novo* mutations in the *PTCH1* gene were found in 9 of 13 (69%) BCCs with 9q LoH. Point mutations of both the substitution missense and substitution nonsense types have been described throughout the length of the PTCH1 protein.

The second important gene in BCCgenesis is *TP53* (Shea *et al.*, 1992; Ziegler *et al.*, 1993). *TP53* codes for the P53 protein that maintains genomic stability by regulating the cell cycle, activating DNA repair, and inducing apoptosis. *TP53* is mutated in anywhere from 30 to 70% of basal cell tumors (Lacour, 2002; Reifenberger *et al.*, 2005; Tang, 2011). Thus, mutations in both the SHH/Patched/SMO pathway and TP53 have long been known to play a major role in basal cell carcinogenesis, the former in both the inherited (Gorlin's syndrome) and sporadic forms of the disease.

In contrast to the aforementioned directed approaches that identified the cancer genes *PTCH1* and *TP53*, recent advances in sequencing technology have permitted unbiased

[1]*Division of Dermatology, Carilion Clinic and Virginia Tech Carilion School of Medicine, Roanoke, Virginia, USA and* [2]*Division of Dermatology, Harbor-UCLA Medical Center, Torrance, California, USA*

Correspondence: Michael S. Kolodney, Division of Dermatology, Carilion Clinic–Riverside Facility, 1 Riverside Circle, Suite 300D, Roanoke, Virginia 24016, USA. E-mail: mskolodney@carilionclinic.org

Abbreviations: BCC, basal cell carcinoma; LoH, loss of heterozygosity; PTCH1, Patched 1; SCC, squamous cell carcinoma; SHH, sonic hedgehog; SMO, Smoothened

Received 15 January 2013; revised 28 May 2013; accepted 28 May 2013; accepted article preview online 17 June 2013; published online 25 July 2013

Table 1. UV signature mutation spectrum in all samples

Mutation type	C>T transitions at dipyrimidinic sites across all samples (YC:RG>YT: RA) except tandem base substitutions (CC:GG>TT:AA)	Tandem base substitutions at CC sites across all samples (CC:GG>TT:AA), each counted as 2	Total "UV Signature" (Ikehata and Ono, 2011)
Number	18,212	2,440	20,652
Mean proportion of coding mutations	66.7%	8.9%	***75.7%***

G>T (oxidative) and T>G (UVA fingerprint) are not included here because of their uncertain UV origins.

approaches to the discovery of cancer mutations. Directed approaches characterize the prevalence of mutations in known oncogenic genes in a cancer of interest. Unbiased approaches sequence a substantial portion of the tumor's genome rather than specific genes of particular interest. These tumor genetic sequences are compared with sequences from normal cells, of the same individual, to identify cancer-specific mutations. One such unbiased approach is whole-exome sequencing. In this method, the genome is enriched for exons—that portion of the genomic code that exits the nucleus for participation in protein translation. Because such studies examine the entire set of protein coding human genes, these approaches have a greater power to detect novel genes involved in tumorigenesis compared with those that target only a small subset of the human exome. Application of this unbiased approach has identified novel mutations in most common cancers including cutaneous squamous cell carcinoma (SCC) and cutaneous malignant melanoma (Durinck *et al.*, 2011; Wei *et al.*, 2011; Berger *et al.*, 2012; Hodis *et al.*, 2012). However, to our knowledge, a similar effort has not been put into the study of BCCs that are more prevalent than other cutaneous malignancies. We have applied whole-exome sequencing to sporadic BCCs in an attempt to better understand the mutational landscape of the single most common cancer worldwide.

RESULTS

Detection of mutations in protein coding DNA

We performed whole-exome sequencing on paired tumor and normal DNA specimens obtained from 12 tumors (10 patients) with histologically confirmed BCC. In both tumor and normal, 65-fold mean coverage was achieved with 89.2% of bases on average covered at least 15-fold.

A total of 27,286 mutations were uncovered in the "protein coding" DNA regions spanning ~30 million base pairs per human genome (Ng *et al.*, 2009) (to be contrasted with "exonic" regions; nearly half of all "exonic" regions are not "protein coding" because the human genome has massive lengths of 5′ and 3′ untranslated regions) across all samples, giving a rate of 75.8 mutations per Mb of coding DNA. This figure is the highest reported in any type of human cancer (those arising in patients with nucleotide excision repair defects may be a possible exception) and more than twice as much as the previously reported high with cutaneous SCC, which is a distant second at 33.3 per Mb of coding DNA (Durinck *et al.*, 2011). Several other types of human cancers have been shown to have mutation rates that are much lower than these. The interested reader is referred to Table 1 in

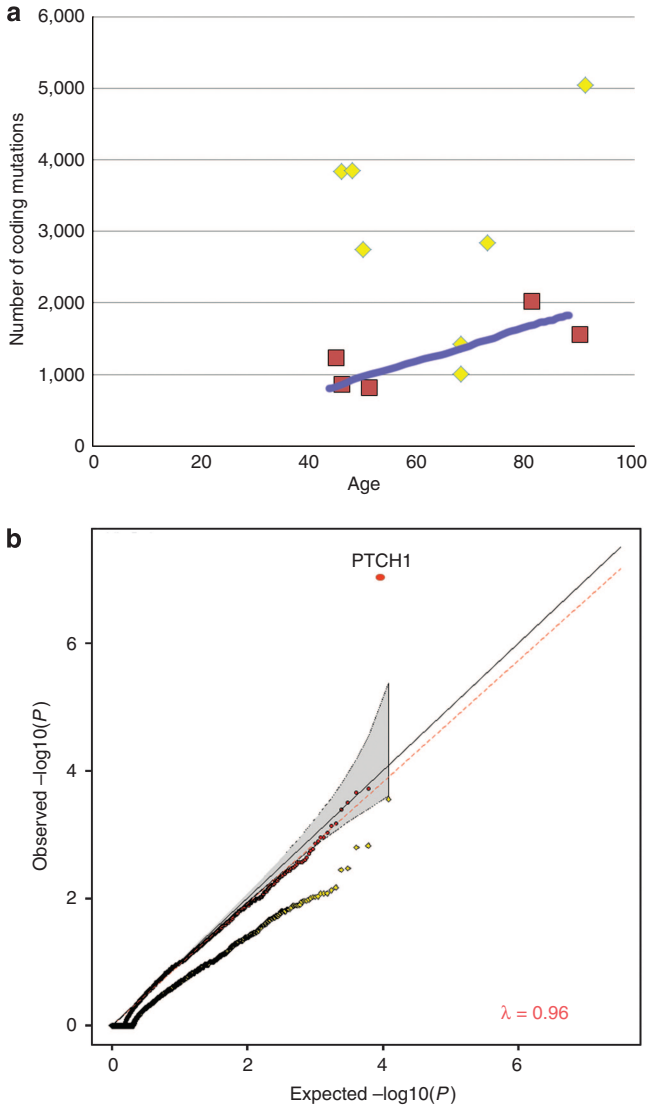

Figure 1. (**a**) Number of coding mutations versus age and photo exposure; each data point represents one tumor. There is a slight trend (blue line) of increasing mutation rates with age in tumors from "protected" (intermittently sun-exposed) parts. Mutations rates are higher in chronically sun-exposed versus intermittently sun-exposed parts (yellow vs. red data points). (**b**) Q-Q plot of functional mutation burden test and synonymous mutation burden test across all genes with at least 1 mutation in the set of 12 tumors. Gray-shaded area represents 95% confidence intervals for expected *P*-values. Red points: functional mutation burden; yellow points: synonymous mutation burden.

Greenman *et al.* (2007) who use direct PCR sequencing of a subset of the protein coding regions to calculate mutation rates on multiple tumor types. As expected, a majority of point