Universidad Autónoma de Madrid
Facultad de Ciencias
Departamento de Biología Molecular

TESIS DOCTORAL

Use of the Sleeping Beauty transposon system to identify new genes involved in skin and breast cancer

Rita Quintana Portilla
Madrid, 2012

Universidad Autónoma de Madrid
Facultad de Ciencias
Departamento de Biología Molecular

TESIS DOCTORAL

Use of the Sleeping Beauty transposon system to identify new genes involved in skin and breast cancer

Memoria presentada por **Rita Quintana Portilla** para optar al Grado de Doctora por la Universidad Autónoma de Madrid

El trabajo experimental presentado en esta memoria ha sido realizado en la Unidad de Oncología Molecular del Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) bajo la dirección de los Doctores Ángel Ramírez Merino y Manuel Navarro Espinel.

Dr. Ángel Ramírez Merino **Dr. Manuel Navarro Espinel**

Rita Quintana Portilla
Madrid, 2012

SUMMARY

New high throughput technologies have made it possible to screen complete cancer genomes for identification of mutations associated to cancer. These analyses have pointed out that there are only a few commonly mutated genes and many infrequently mutated genes, highlighting the heterogeneity in the genes that cause cancer. The validation of the possible role for each of these genes in cancer is a difficult task, considering the existence of driver mutations and many "passenger" mutations in cancer genomes. To overcome these difficulties, comparative genomics has proved to be a useful tool when looking for candidate cancer genes that are mutated in tumors that arise in model mice. Transposons represent important genetic tools for functional genomic screening and recently the Sleeping Beauty transposon system, which can act as an insertional mutagen, has been developed to mutate and identify novel cancer genes in transgenic mice. The major aim of this project is to identify novel cancer genes involved in the development of two tumor types that are incompletely characterized at genetic level: non-melanoma skin cancer and breast cancer malignancies. For this purpose, we have used the Sleeping Beauty transposon system to identify somatic mutations associated with skin and breast cancers.

We generated transgenic mice carrying multiple copies of the mutagenic transposon T2Onc2 and expressing the SB11 transposase specifically in tissues where keratin K5 is expressed (SB/T2 mice). When these mice were subjected to several skin carcinogenesis regimes, they developed more tumors than mice lacking active transposons in the skin. Moreover, these tumors showed worse histological characteristics than those obtained in control mice, presumably as a result of the mutagenic effect of transposon mobilization. Genetic analysis of the transposon insertion sites in the tumors identified 126 genes recurrently mutated in different tumor samples, which may represent candidate cancer genes. When the expression levels of a subset of the most transposon-mutated genes were analyzed in skin human samples, we detected alterations in the expression levels of some of these genes in human tumors. Our results show that inactivating mutations in Notch1, Nsd1 among others, may have an important role in skin carcinogenesis. We also generated heterozygous p53+/- mice bearing multiple copies of the T2Onc2 transposon and the SB11 transposase. These mice developed mammary gland carcinomas with higher incidence and shorter latency period than control mice that did not undergo transposition. High throughput sequencing analysis of the transposon insertion sites within the tumors identified 53 candidate cancer genes that were commonly mutated in different transposon/transposase tumor samples. Immunohistrochemical analyses revealed that *NF1* and *RASA1* (two of the most frequently transposon-mutated genes in our animal model) were expressed at a reduced level in human breast tumor samples. Therefore, this study shows that NF1 and RASA1 may be important for breast cancer development.

RESUMEN

Las nuevas técnicas de secuenciación masiva han posibilitado el análisis genómico de tumores para identificar mutaciones asociadas a cáncer. Estos análisis han averiguado que hay un número pequeño de genes mutados frecuentemente y gran número de genes mutados infrecuentemente, poniendo de manifiesto la heterogeneidad de los genes que causan cáncer. La validación del posible papel de cada uno de estos genes en cáncer es una tarea difícil, ya que hay mutaciones conductoras y pasajeras en los genomas de cáncer. Para vencer estas dificultades, la genómica comparativa ha demostrado ser una herramienta muy útil para la búsqueda de genes importantes en el desarrollo del cáncer, a través de la determinación de los genes mutados en tumores originados en modelos de ratón. Por otra parte, los transposones representan herramientas genéticas muy importantes para el análisis genómico; recientemente se ha desarrollado el sistema de transposición "Sleeping Beauty" (SB) para mutar e identificar nuevos genes implicados en cáncer en modelos animales. Finalmente, tanto el cáncer de piel (no melanoma) como el cáncer de mama están incompletamente caracterizados a nivel genético. El principal objetivo de este trabajo es la identificación de mutaciones somáticas asociadas a estos dos tipos de cáncer mediante el uso del sistema de transposición SB en ratones transgénicos.

Hemos generado ratones transgénicos portadores de múltiples copias de un transposón mutagénico y de la transposasa necesaria para su movilización específicamente en tejidos donde se expresa la queratina K5 (ratones SB/T2). Cuando estos ratones fueron sometidos a varios protocolos de carcinogénesis de piel, desarrollaron más tumores que ratones sin transposones activos en piel. Además, estos tumores presentaron peores características histopatológicas que los obtenidos en animales control. El análisis de los sitios de inserción del transposón en los tumores identificó 126 genes mutados de manera recurrente. Estos genes son candidatos a desempeñar un papel importante en el desarrollo tumoral. Nuestros resultados muestran que mutaciones que inactivan NOTCH1 y NSD1, entre otros genes, puede ser importante para la carcinogénesis de piel.

También hemos generado ratones SB/T2 en un fondo genético heterocigoto para *Trp53*. Estos ratones desarrollaron carcinomas de glándula mamaria con mayor incidencia y menor latencia que ratones sin transposición activa. El análisis de los sitios de inserción en los tumores mamarios por secuenciación masiva identificó 53 genes candidatos de cáncer que estaban mutados de manera recurrente en diferentes tumores. Estudios inmunohistoquímicos pusieron de manifiesto que dos de los genes más mutados por el transposón en nuestro modelo animal (en concreto NF1 y RASA1) se expresan a menor nivel en muestras humanas de cáncer de mama que en tejido no tumoral. Estos resultados indican que NF1 y RASA1 pueden ser importantes para el desarrollo del cáncer de mama.

ABBREVIATIONS

- 454/Roche — Genome Sequence FLX platform
- AK — Actinic keratosis
- B6D2 — C57BL/6J X DBA/2J
- BCC — Basal cell carcinoma
- BP — Biological process
- BSA — Bovine serum albumin
- CAG — CMV enhancer/ chicken b-actin promoter
- CC — Cellular component
- cDNA — Complementary deoxyribunocleic acid
- CISs — Common insertion sites
- COSMIC — Catalogue of somatic mutations in cancer
- DAPI — 4',6-diamidino-2-phenylindole
- dATP — deoxyadenosine tryphosphate
- DAVID — Database for annotation, visualization and integrated discovery
- DCIS — Ductal carcinoma in situ
- DMBA — 7,12-dimethylbenz[a]anthracene
- DNA — Deoxyribunocleic acid
- dNTP — deoxynucleotide tryphosphate
- DTT — Dithiothreitol
- EDTA — Ethylenediaminetetraacetic acid
- EGTA — Ethylene glycol tetraacetic acid
- EMT — Epithelial-mesenchymal transition
- En2 — Engrailed 2 splice acceptor
- ENU — N-ethyl-N-nitrosourea
- ER — Estrogen receptor
- FITC — Fluorescein Isothiocyanate
- FVB/J — FVB genetic background
- GAPs — GTPase activating proteins
- gCIS — gene-centric CIS analysis
- GTPase — GTP phosphatase
- HEPES — N-2-Hydroxyethylpiperazine-N'-2-Ethanesulfonic Acid
- HH — Hedgehog
- hnRNP — Heterogeneous nuclear ribonucleoprotein

- HPV Human papillomavirus
- ID Identifier
- IF Immunoflourescence
- IHC Immunohistochemistry
- IR Inverted repeat
- IRL Inverted repeat left
- IRR Inverted repeat right
- K14 Keratin 14
- K5 Keratin 5
- K8 Keratin 8
- KA Kerathoacantomas
- KEGG Kyoto Encyclopedia of Genes and Genomes
- LCIS Lobular carcinoma in situ
- LM-PCR Ligation-mediated PCR
- LTRs Long terminal repeats
- MF Molecular function
- MMTV Mouse mammary tumor virus
- MOPS 3-(n-Morpholino)Propanesulfonic acid
- mRNA Messenger ribonucleic acid
- MSCV LTR Long terminal repeat of the murine stem cell virus
- MuLV Murine leukemia virus
- NEB New England Biolabs
- NMSC Non-melanoma skin cancer
- pA Polyadenylation
- PAGE Polyacrilamide Gel Elecrophoresis
- PBS Phosphate buffered saline
- PCR Polymerase Chain Reaction
- PR Progesterone receptor
- RNA Ribonucleic acid
- RNase Ribonuclease
- Rosa26-SB11 SB11 inserted into the Rosa 26 locus
- SA Splice acceptor
- SB transposon Sleeping Beauty transposon
- SB K5-SB11 transgenic mice
- SB/p53+/- K5-SB11;p53+/- transgenic mice
- SB/T2 K5-SB11;T2/Onc2 transgenic mice
- SB/T2/p53+/- K5-SB11;T2/Onc2;p53+/- transgenic mice

- SB/T2/Tg.AC K5-SB11;T2/Onc2;Tg.AC transgenic mice
- SB/Tg.AC K5-SB11;Tg.AC transgenic mice
- SCC Squamous cell carcinoma
- SD Splice donor site
- SDS Sodium Dodecyl Sulfate
- SHH Sonic hedgehog
- SSC Saline Sodium Citrate
- T2 T2/Onc2 transgenic mice
- T2/p53+/- T2/Onc2;p53+/- transgenic mice
- T2/Tg.AC T2Onc2;Tg.AC transgenic mice
- TA Thymine Adenine dinucleotides
- TAE Tris-acetate EDTA buffer
- TBE Tris-borate EDTA
- TBS Tris-buffered saline
- TDLUs Terminal ductal lobular units
- TE Tris-EDTA
- Tg Transgenic
- TPA 12-O-tetradecanoylphorbol-13-acetate
- UV Ultraviolet radiation
- WAP Whey acidic protein
- Wt Wild-type

INTRODUCTION

1. NON-MELANOMA SKIN CANCER

1.1 The biology of non-melanoma skin cancer

The skin epidermis is a stratified epithelium composed principally of keratinocytes, whose proliferation and differentiation must be tightly regulated and coordinated in order to maintain skin homeostasis and determine effective response to injury. Failures in the regulation of these processes precipitate skin disorders, such as skin cancer. Cells that reside at the base of the epidermis are called basal keratinocytes and are less differentiated and possess proliferative potential. As basal keratinocytes divide, their progeny detach from the underlying basal lamina and migrate vertically. Thus, they become postmitotic and committed to an irreversible differentiation program, culminating in enucleation, internal degradation and transformation into squames that constitute the most superficial layer or stratum corneum. Since the outermost squames are continuously shed from the surface of the skin, the basal cells of the epidermis maintain a steady proliferative rate to compensate from this loss (**Figure 1**)

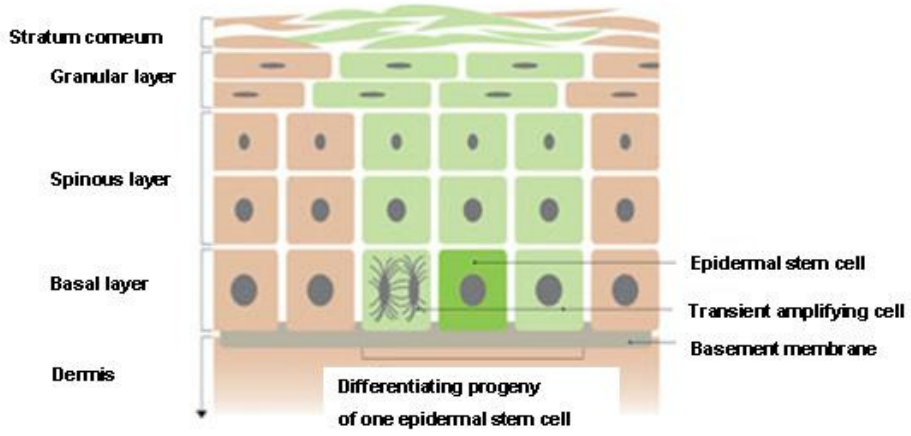

Figure 1. Epidermal stem cells are located in the basal layer of the epidermis. These stem cells give rise to transient amplifying cells that migrate to the surface undergoing a differentiation process

The process of skin tumor formation follows 3 steps: tumor initiation, tumor promotion and tumor progression. Tumor initiation involves crucial gene mutations in proto-oncogenes and/or tumor suppressor genes in skin cells. Tumor promotion consists of clonal expansion of these cells that will give rise to a benign tumor. Tumor progression involves the conversion of the benign tumor into an invasive and potentially metastatic malignant tumor (**Figure 1**).

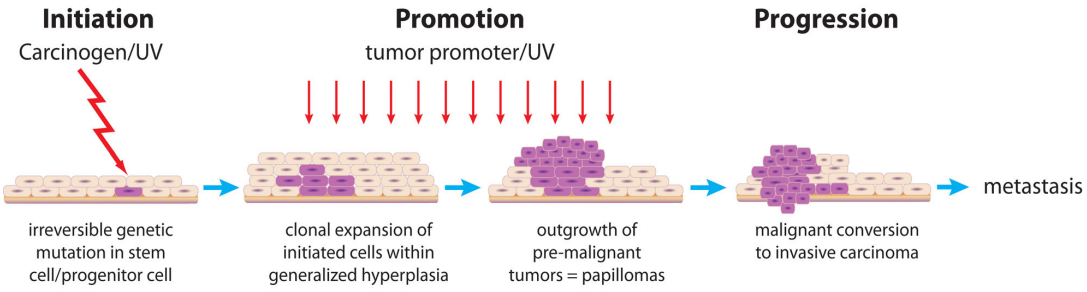

Figure 2. Schematic representation of skin tumor development

Non-melanoma skin cancer (NMSC) is the most common form of human cancer, with incidence rates dramatically increasing over the last decade, probably as a result of increased sun exposure and the depletion of the ozone layer. The two more frequent types of cancer in NMSC are basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). Approximately, 80% of the NMSC cancers are BCCs and 20% correspond to SCCs. BCCs are keratinocyte tumors characterized by local invasion and contiguous spread that cause tissue destruction, but rarely metastasize. Most of BCCs occur sporadically, but they also develop as part of heritable syndromes like Gorlin syndrome, linked to mutations in the *PTCH1* gene (90). They classically appear as lesions on the sun-exposed areas of skin, being much more frequent in fair-skinned individuals of European ancestry. Histologically, BCC cells resemble to basal cells of the interfollicular epidermis, the hair follicle or the sebaceous gland and it is thought that BCC arise from the basal cells specifically of the hair follicle. SCCs tumors also arise from cutaneous keratinocytes. However, they are biologically aggressive tumors that after local invasion and tissue destruction may metastasize (1-12.5%) to lymph nodes and can cause death if not treated properly. While BCCs are believed to develop *de novo*, SCCs development follows a mutational multistep process, characterized by well-defined precancerous skin lesions that can lead to actinic keratosis (AK) or kerathocantomas (KA). AKs consist of a focal dysplasia of epidermal keratinocytes that is followed by *carcinoma in situ* or Bowen's disease, that represents a preinvasive stage of skin SCCs. KAs are benign cutaneous squamous neoplasias arising preferentially on sun-exposed skin that have been also classified as premalignant SCC lesions. The incidence of SCC is greatly increased in immunosuppressed patients and human papilloma virus (HPV) infected individuals. A number of heritable conditions are associated with an increased risk of SCC, including xeroderma pigmentosum, recessive dystrophic epidermolysis bullosa and dyskeratosis congenita but contrary to what happens to BCC, the genetic defects in these syndromes are not specific for SCC. They are believed to arise as a secondary result of increased epidermal mutagenesis.

Although some specific therapies have been developed to treat NMSC, surgery still remains as the primary therapeutic modality.

1.2 Mutated genes and deregulated pathways in non melanoma skin cancer

Cumulative exposure to ultraviolet radiation (UV) is considered as the most common risk factor of NMSC, but other factors such as immunological status, genetic predisposition and infection by HPV may also be involved. UV light induces DNA damage in keratinocytes, and it can mutate genes essential for control and surveillance in the skin epidermis. Damaged DNA cells are repaired and if this it not possible, they are eliminated by apoptosis (sunburn cells), a process that occurs under the control of the *P53* gene. *P53* gene mutations by UV radiation leads to genome-wide instability, chromosomal amplifications and wide allelic imbalance; being the starting point of SCCs and some BCCs. Several other gene pathways have been correlated to NMSC development and progression. So, mutations in several genes of the *SHH* pathway are known to be a main cause of BCCs; the Ras pathway is altered in SCCs and BCCs; and deregulation of *MYC*, *CDKN2A* and *NF*-κB has also been reported in this type of cancers.

1.2.1 P53 mutations in NMSC

The *TP53* tumor suppressor gene encodes the P53 nuclear transcription factor that plays a central role in the decision between DNA repair or apoptosis following UV radiation-mediated damage. It arrests cell division at G1 phase to allow DNA repair by promoting p21-mediated inhibition of cyclin-dependent kinases complexes. Cell apoptosis is induced through p53-induced expression of genes like *NOXA*, *PUMA* or *BAX* if the genomic damage cannot be repaired (**Figure 3**). Additionally, an antioxidant function for this gene has also been proposed, protecting the cell against oxidative stress

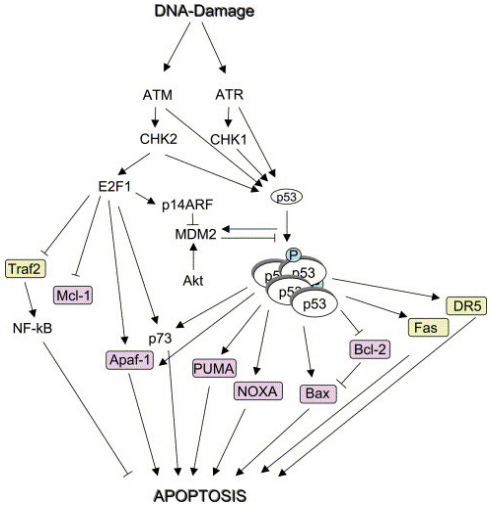

Figure **3**. Schematic representation of the P53-mediated response after DNA damage. Activation of p53 results in transcriptional activation of a variety of genes including, NOXA, PUMA, BAX and APAF-1 that contribute to apoptosis. (Modified from Crighton (36))

Mutational inactivation of the *P53* gene allows damaged or incompletely repaired cells to progress through the cell cycle and to pass their damaged genomes to daughter cells. Mutations

in *P53* are considered to be the earliest event in NMSC and 50% of these cancers are p53-mutated. Many *P53* mutations found in these tumors consist of C to T or CC to TT transitions, consistent with an UV radiation signature (23). These mutations do not appear at random, showing a pattern of hot spots different from that of internal malignancies (222). These data support the role of UV light as the main cause for *P53* gene inactivation in NMSC. Although mutations in *P53* have been found in both BCC and SCC, it has been proposed an important role of *P53* mutation in the pathogenesis of SCCs. According to experimental data, more than 70% of preinvasive SCCs display p53 mutations (28), implicating that p53 mutation is an early event in SCC development.

However, it has been found that *Trp53* knockout mice display a surprising resistance to skin carcinogenesis induced by overexpression of certain oncogenes (67). In addition, results of experimental approaches to alter P53 function vary according to genetic deletion versus overexpression of well-characterized mutant P53 proteins. Thus, it is still necessary to perform further studies to understand the functional consequence of p53 mutations in NMSC development.

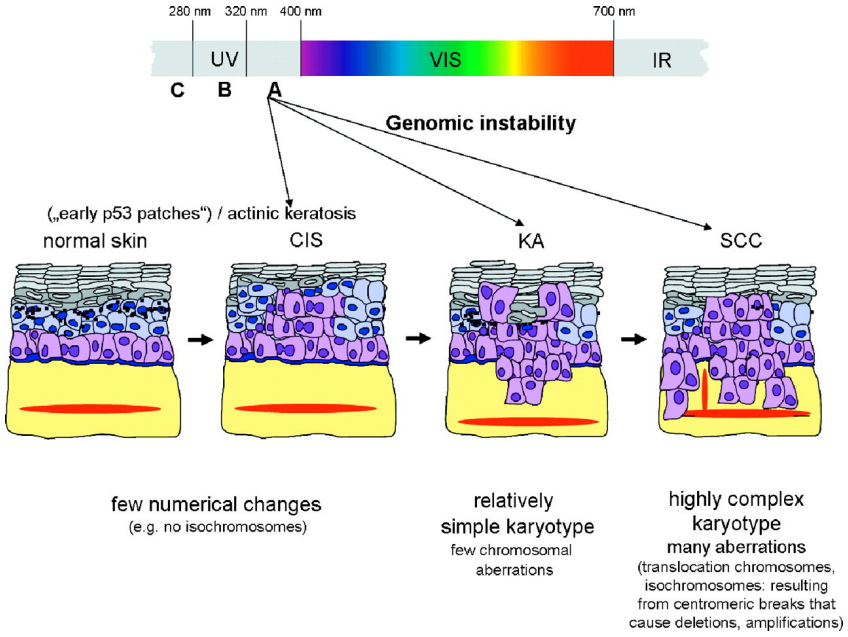

Figure 4. Schematic representation of UV-induced skin carcinogenesis. Both UV-B and UV-A can contribute to tumor development.by direcly inducing DNA damage, such as P53 mutations or indirectly generating genomic instability. (Modified from Boukamp,(24)) CIS (Carcinoma in situ)

1.2.2 Role of the Hedgehog signaling pathway in NMSC

The Hedgehog (HH) signaling pathway is involved in the regulation of embryo growth and patterning in mammalian development, and has also been implicated in the maintenance of stem or progenitor cells in many adult tissues. Mutations in components of this pathway are involved

in the development of a broad list of cancers including medulloblastoma, breast carcinoma, meningioma, colon carcinoma, pancreatic and oesophageal adenocarcinomas and small cell lung cancer.

The *PTCH1* gene encodes a transmembrane protein that acts as a receptor for the Hedgehog extracellular ligands, being sonic hedgehog (SHH) the best characterized ligand. The SHH ligand binds to PTCH1 protein to activate the pathway. In the absence of SHH binding, PTCH1 holds the smoothened (SMO) protein, a G protein-coupled receptor, in an inactive state, inhibiting signaling to downstream genes. When SHH ligand binds to PTCH1, inhibition of SMO is released and the signal is transduced through a series of interacting proteins including suppressor of fused (SUFU), leading to the activation of the transcription factors *GLI1*, *GLI2* and *GLI3*. Their target genes include cell cycle regulators, *WNT*, *TGFB* and *PTCH1* and *GLI1*. The upregulation of *PTCH1* provides negative feedback and that of *GLI1* provides positive feedback to the HH signaling pathway (**Figure 5**).

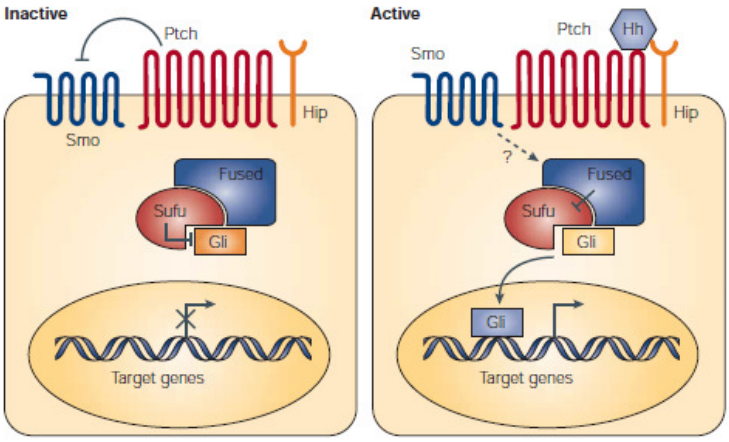

Figure 5. Schematic representation of SHH signaling pathway. When SHH ligand binds to PTCH, SMO becomes active and initates a signaling cascade that results in the activation of GLI transcription factors. GLI proteins interact in a complex that includes Fused (FU) and suppressor of FU (SUFU). (Modified from Pasca di Maliano and Hebrok,(135)**)**

The Gorlin syndrome is an autosomal dominant inherited syndrome in which patients have a marked susceptibility to develop BCCs (70). Family-based linkage studies of kindreds with Gorlin syndrome resulted in the identification of mutations in the *PTCH1* gene. After that, *PTCH1* mutations were also found in sporadic BCCs (9). These findings linked this pathway to BCC development and progression. Approximately, 70% of sporadic BCCs have identifiable mutations in at least one allele of *PTCH1*. Many of these mutations are consistent with an UV signature, indicating an UV-radiation origin. Another 10% of BCCs have activating mutations in *SMO* gene (208). A smaller number of BCCs have mutations in *PTCH2*, the *PTCH1* homologue and in *SUFU*. In addition, It has been found that BCCs overexpress GLI1 and GLI2,

indicating that the HH pathway is constitutively active in BCC (10),(148). However, mutations in *PTCH1*, *SMO* or *SUFU* have not been found in all BCCs, so it seems likely that other genes of the HH pathway maybe mutated in some BCC. Although several experiments have been done to determine genome-wide expression differences in BCC tumor samples versus control skin, there is only a limited understanding of which downstream expression changes are actually crucial for HH-induced BCC carcinogenesis.

Possible downstream mediators of BCC carcinogenesis from HH signaling activation include up-regulation of apopotosis inhibitors, as BCL-2 (147) and cFlip (58); downregulation of apoptosis inducers as the tumor necrosis factor superfamily member FAS (110) and BMI1 (149), as well as up-regulation of *PDGFR* (207). Other downstream HH components are the transcription factors *FOXM1* and *FOXE*. Their overexpression leads to activation of the cell cycle and proliferation. Overexpression of *FOXM*1 has been found in BCC tumors (182). However, it is still unclear which downstream changes are essential for BCC tumorigenesis.

In addition, other signaling pathways are likely to be involved through their interaction with the HH pathway, such as those pathways regulated by EGFR, PI3K/AKT and PKC. Experimental data revealed that EGFR signaling through RAS/RAF/MEK/ERK induces Jun/AP1 activation that acts synergistically with activated forms of GLI1 and GLI2 in epithelial oncogenic transformation (158). Data also show that PI3K and AKT activities are essential for Gli-SHH signaling and that PI3K/AKT signaling potentiates GLI1 activation induced by low levels of SHH (150). It has also been proposed that PI3K/AKT can stabilize GLI2 through inhibition of PKA. These data determine a synergistic relation between the PI3K/AKT and HH pathways.

1.2.3. Wnt signaling

A recent study presents convincing evidence regarding the role of Wnt signaling in BCC carcinogenesis, being downstream of HH signaling in these tumors (211). These authors have reported the need of canonical Wnt/β-catenin signaling for HH-driven tumorigenesis, as Wnt pathway inhibition prevented the formation of tumoral lesions in animal models without affecting HH signaling.

1.2.4. Cell cycle inhibitors

The *CDKN2A* locus encodes $p16^{INK4a}$ and $p14^{ARF}$ cell cycle inhibitors. $p16^{INK4}$ inhibits cell cycle progression in the G1 phase by binding and inhibiting CDK4/6 kinases, that phosphorylate the RB protein. Upon RB phosphorylation, activated E2F promotes induction of $p14^{ARF}$, which sequesters MDM2 preventing degradation of p53. Loss of these proteins leads to the promotion of cell division. Mutations in the *CDKN2A* gene have been found in sporadic SCC (154) and overexpression of *CDK4* and oncogenic *RAS* is sufficient to induce human SCCs (105).

1.2.5. The RAS proto-oncogene

The Ras gene family encode Ras GTPase proteins that regulate proliferation, angiogenesis, apoptosis and cellular morphology and are considered to be among the most widely altered oncoproteins in human neoplasia. The human *H-RAS* proto-oncogene is the most frequently activated member of the Ras family in NMSC. Consequently, activation of Ras/Raf signaling has been implicated in both BCCs and SCCs and cDNA studies have shown an induction of known Ras target genes in human SCC patient tissue versus matched normal skin samples. However, only an overall H-ras mutation frequency of 10/20% has been found in NMSC tumors (114), (29).

1.2.6. The NF-κB pathway

NF-κB proteins are conserved transcription factors implicated in regulating inflammation, morphogenesis, apoptosis, differentiation and proliferation. In contrast to its oncogenic role in other settings, NF-κB acts as an important proliferative safeguard in stratified epithelium. Its blockade is related to hyper-proliferation in both murine SCC (188) and human epidermis (160). Additionally, it has been shown that NF-κB blockade and expression of oncogenic Ras produced a SCC-like neoplasia in normal human epidermal cells (38).

1.3. Mouse models used for the study of human skin cancer

1.3.1. The multistep chemical carcinogenesis skin tumor model in mice

The multistage skin carcinogenesis protocol using DMBA/TPA has been widely used to address important skin carcinogenesis process questions (141). It involves treatment of the dorsal skin of mice with a single dose of the mutagenic agent 7,12-dimethylbenz[a]anthracene (DMBA), followed by weekly applications of the hyperplasic agent 12-O-tetradecanoylphorbol-13-acetate (TPA). DMBA causes mutations in the *H-ras* oncogene (141), (12). TPA is a tumor promoter that activates PKC (186). Upon activation by TPA, PKC sets in motion pathways leading to the upregulation and activation of transcription factors such as c-fos and c-jun (AP1) that play an important role in initiating cell proliferation. During papilloma promotion chromosomal abnormalities are also frequent (4, 34).

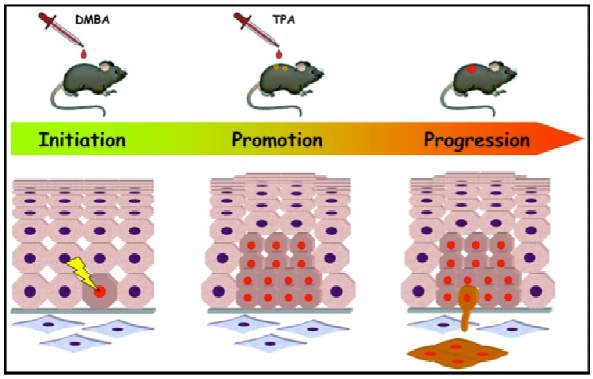

Figure 6. Schematic representation of the DMBA/TPA protocol. DMBA causes intial mutations and TPA causes an hyperproliferation of the initially mutated cells that can lead to papilloma formation. Papillomas can then originate SCC-like tumors and metastasize

This process gives rise to papillomas, which are benign neoplastic lesions consisting of hyperplastic keratinocytes and supporting stromal cells. Although many papillomas regress, a small percentage of papillomas progresses to malignant, invasive squamous cell carcinomas (73). Carcinomas invade through the basement membrane and are characterized by loss of differentiation, increased mitotic activity, and nuclear atypia. A small fraction of these carcinomas can metastasize to distant places.

The conversion to carcinomas is characterized by loss of heterozigosity and mutation of p19/Arf and $p16^{Ink4a}$ or *Trp53* (99), (94), and increases in mutant *H-ras* via gene amplification (26). In fact, reduction of Trp53 gene dosage enhances malignant progression of chemically induced skin tumors (93). These aggressive carcinomas also present inducted overexpression of *Snail* and *Slug* genes involved in epithelial mesenchymal transition (EMT) (22), (30).

1.3.2. The Tg.AC transgenic mouse

These mice represent an improvement to the DMBA/TPA protocol, since they do not need to be initiated. These mice carry a transgene consisting of an oncogenic *H-ras* coding region flanked in its 5' region by a mouse ζ-globin promoter and in 3' by an SV-40 polyadenylation sequence. Located on chromosome 11, the transgene is transcriptionally silent until activated by hyperproliferative stimuli, including TPA, UV light, or full thickness wounding. Expression of the transgene is an early event that drives cellular proliferation resulting in clonal expansion and tumor formation. When these mice are treated topically with TPA, they develop multiple papillomas and some of them can progress to malignant spindle or squamous cell carcinomas (107), (169)

1.3.3. Mouse models of BCC carcinogenesis

The DMBA/TPA and Tg.AC classical skin cancer models have been very useful to study genes related to SCC tumors, but these models are not able to produce BCC-like tumors. Therefore,

several mouse models were engineered carrying inactivating mutations in genes that belong to the HH signaling pathway.

Ptch1+/- mice are prone to develop BCCs after receiving mutagenic ionizing or UV radiation(8),(120). Overexpression of *SHH* and active *SMO*, *GLI1* or *GLI2* genes promote development of BCC-like tumors (9), (134), (66). These experiments have highlighted the importance of alterations in the HH signaling in BCC development. Additionally, it has been seen that conditional *Trp53* loss in K14 or K15-expressing keratinocytes markedly enhances HH-driven tumorigenesis in *Ptch1* +/- mice (199), reinforcing the role of p53 in BCC carcinogenesis. Additionally, results of pharmacological interventions in the Ptch1+/- mouse seem to correlate well with the results of the same intervention in humans (180).

2. BREAST CANCER

2.1 The biology of breast cancer

Breast cancer is the most commonly diagnosed cancer and the leading cause of cancer death among women worldwide, accounting for 23% of the total new cancer cases. More than 1.4 million new cases of breast cancer occurred among women worldwide in 2008 (25). The mammary gland consists of a branching ductal system that ends in terminal ducts with their associated acinar structures, termed the terminal ductal lobular units (TDLUs), together with interlobular fat and fibrous tissue. TDLUs comprise two types of epithelial cells: inner secretory luminal cells (ductal and alveolar subtypes) and outer contractive myoepithelial/basal cells. The stroma is composed of adypocites, fibroblasts and different immune cells.

The initiation of breast cancer is caused by transforming events, including genetic and epigenetic changes in a single cell. Subsequent tumor progression is driven by the accumulation of additional genetic changes combined with clonal expansion and selection. The process involves hyperproliferation of ductal cells that will lead to ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS), followed by evolution into an invasive carcinoma and metastasis. Transformation from carcinoma in situ to invasive carcinoma comprises the disappearance of the myoepithelial cell layer as an organized entity, so ductal cells from inside lobules or ducts can invade nearby breast tissue. Metastasis process involves spread of mammary gland tumor cells through blood and lymphatic vessels to distant organs, more frequently lungs, bones and liver (**Figure 7**)

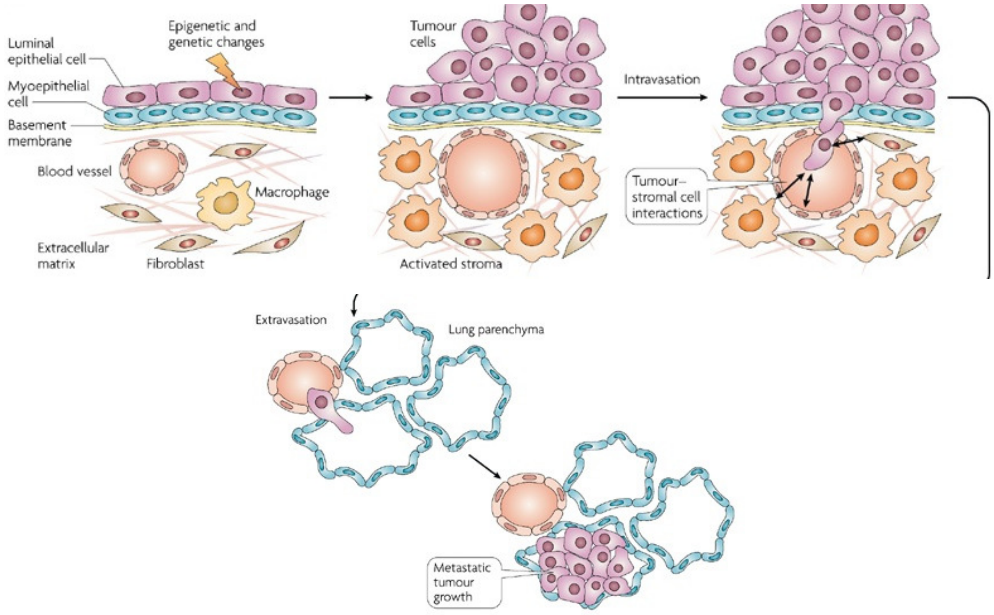

Figure 7. Schematic representation of breast cancer progresión. Epigenetic and genetic changes and interactions with the microenvironment cause the transformation of normal breast epithelial cells to tumor cells that can lead to metastatis by extravasation to distant organs. (Modified from Polyak (137))

Five major molecular subtypes of ductal breast cancer have been determined by gene expression profiling of large sets of tumors: Luminal A, Luminal B, HER2^{+}, Basal like ("triple negative") and Claudin-low; each subtype has a different rate of incidence, survival and response to treatment (183). Luminal A represents the most frequent form of breast cancer and has the best prognosis, while the basal-like has the worst outcome being frequently associated with metastasis. They have been classified according to expression of progesterone receptor (PR), estrogen receptor (ER) and Erbb2 (HER2) expression status in the tumors (**Figure 8**)

Luminal A	ER+ and/or PR+, Ki67-
Luminal B	ER+ and/or PR+; Ki67+
Her2	HER2+
Basal-like	ER- and PR- and HER2 -
Claudin-low	ER- and PR- and HER2 -; stem cell-like

Figure 8. Molecular classification of breast tumors

2.1.1. Keratin distribution in mammary cell lineages and mammary epithelial stem cells

The mammary gland development is characterized by different keratin expression along the process. Each mammary cell lineage shows a specific keratin expression pattern. In adult tissue, luminal cells (ductal and alvelolar) express preferentially keratin 8 (K8) and keratin 18 (K18); and myoephitelial cells are characterized by expression of keratins K5 and K14. However,

between embryonic days E15.5 and E18.5 in the mouse, nearly all epithelial cells express both K5 and K14.(177) Experiments in vitro have shown that this keratin expression pattern characteristic of mouse mammary cell lineage is also conserved in primary human mammary epithelial cells (177). The mammary epithelium suffers deep remodeling processes in each menstrual cycle and in pregnancy, lactation and involution. This has lead to the proposal of the existence of adult mammary epithelial stem cells. According to this, implantation experiments in the mouse have revealed the presence of progenitor cells that when transplanted into a mouse cleared fat pad can regenerate the mammary gland (102), (174), (161). Breast epithelial stem cells are bipotential stem cells able to self-renew and able to produce luminal and myoepitelial progenitors (**Figure 9**).

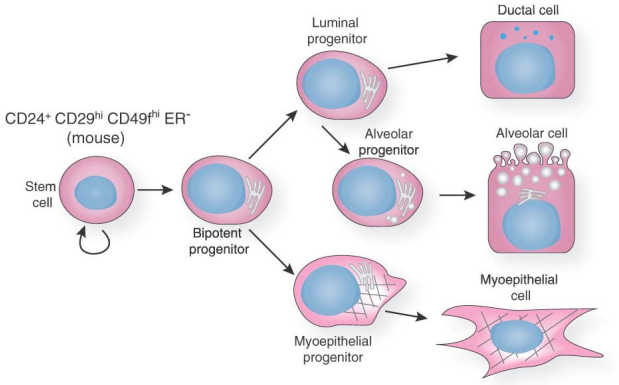

Figure 9. Schematic representation of the epithelial cell hierarchy within the mammary gland. A stem cell divides to generate a bipotent progenitor that originates both luminal and myoepithelial progenitor; these differentiate into myoepithelial cells. Luminal progenitors differentiate into cells of ductal or alveolar lineales. (Modified from Macias (119))

Besides its normal function in tissue development, the importance of mammary epithelial stem cells has been addressed in the development of breast cancer. It has been proposed that breast stem cells would be the tumor-initiating cells since they exist quiescently over long periods of time and could, therefore, accumulate mutations over the life-span of an organism and give rise to tumors when stimulated to proliferate. In addition, it has been seen that around 40% of breast tumors recur 10 years after removal of the primary tumor. This could be due to the existence of cells with a long life span and self-renewal capacity, similar to stem cells behavior.

2.2. Genes mutated in breast cancer

Somatic breast cancer may arise from a combination of multiple acquired genetic events that include gain-of-function mutations in proto-oncogenes that stimulate cell growth, division and survival; and loss-of-function mutations in tumor suppressor genes that normally prevent abnormal cellular growth and promote DNA repair and cell cycle checkpoint activation.

Epigenetic changes also contribute to deregulate the expression of these genes. Only 25% of breast cancer cases are due to inherited gene mutations.

Breast cancer genome analysis indicates that there are only a few genes that are frequently mutated but many that are infrequently mutated. These data provides an explanation for the observed breast cancer molecular heterogeneity. *ERBB2*, *PI3KCA*, *MYC* and *CCND1* oncogenes are frequently deregulated in breast cancer. Among tumor suppressor genes, *BRCA1, BRCA2*, *TP53* and *PTEN* are the most mutated genes.

2.2.1. Role of oncogenes

PI3KCA is the gene most mutated somatically in human breast cancer samples according to the Catalogue of Somatic Mutations in Cancer (COSMIC). Alterations in upstream components of the PI3K pathway, such as receptor tyrosine kinases, and downstream components, such as AKT, that regulate control of cell proliferation, cell survival and metabolism, are frequent in breast cancers. It has been proved that mutations in *PI3KCA* lead to a constitutive activation of *AKT* in human mammary gland epithelial cells (83). Direct evidence supporting a role for the *AKT* family of genes in breast cancer has been also obtained using transgenic mice carrying activating forms of this gene (82).

ERBB2 (known also as *HER2* or *HER2/neu*) is a member of the EGFR family of receptor tyrosine kinases (RTKs). ERBB2 expression levels are elevated in 20/30% of human breast cancers because of genomic amplification of the *ERBB2* proto-oncogene (166). Overexpression of ERBB2 results in aggressive tumors, being frequent in human ductal carcinomas and rarely observed in benign breast hyperplasias. In addition, elevated expression of *ERBB2* is usually associated to tamoxifen resistance of the primary tumor. Studies using transgenic mice carrying the *Erbb2* gene under the transcriptional control of the MMTV promoter or of the *Erbb2* promoter support the role of this gene in breast cancer (68), (6).

MYC encodes a nuclear phosphoprotein involved in cellular proliferation, differentiation and apoptosis. It is amplified and overexpressed in 15-25% of breast tumors (126). Experiments using transgenic mice overexpressing *Myc* under MMTV or WAP promoters showed that these mice developed tumors, but since tumor incidence was low (108) and WAP-*Myc* animals required pregnancy to develop tumoral lesions (159), it was established that additional mutations in other oncogenes were necessary for the induction of carcinogenesis.

CCND1 encodes the protein Cyclin D1, a cell cycle regulator that controls the progression from G1 to S phase. This protein is overexpressed in 40/50 % of invasive breast tumors and its gene is amplified in 10/20% of the cases (172). Lobular carcinoma universally overexpress Cyclin D1, while overexpression in ductal carcinoma appears to be restricted only to ER+ cases (178). Overexpression of this gene under the control of the MMTV promoter in transgenic mice

resulted in mammary adenocarcinomas. Cyclin D1 overexpression is also essential for the development of mammary cancers induced by *ErbB2* and *Ras* expression (214).

2.2.2. Role of tumor suppressor genes

P53 is the second most somatically mutated gene in human breast cancer samples according to COSMIC. It is estimated that 20 to 30% of breast tumors carry mutations in *P53* (76). Breast tumor progression seems to be associated with mutant *P53*, since a higher frequency of *P53* mutations is seen in patients with advanced disease. The role of *P53* mutations in breast cancer is further supported by the fact that Li-Fraumeni syndrome patients, carrying germline *P53* mutations, are predisposed to develop breast cancer at a relatively early age (190) .

PTEN antagonizes the PI3K pathway. Inherited mutations of *PTEN*, such as those in Cowden syndrome, increase the risk of breast cancer (127). In addition, loss of *PTEN* has been related to a worst outcome in breast cancer patients, correlating with metastasis and disease associated death (42). Conditional ablation of *PTEN* gene in the mammary epithelium in transgenic mice accelerated mammary tumor progression (156).

The BRCA1 gene encodes a nuclear phosphoprotein that plays a role in maintaining genomic instability, acting as a tumor suppressor. The encoded protein combines with other tumor suppressors, DNA damage sensors, and signal transducers to form a large multi-subunit protein complex known as the BRCA1-associated genome surveillance complex (BASC). This protein plays a role in transcription, DNA repair of double-stranded breaks, and recombination (61). Linkage analysis of families with multiple cases of breast cancer linked *BRCA1* to hereditary breast cancer. *BRCA1* mutations account for 5% of all breast cancer cases occurring in women under 40 years but rises up to 90% in families with four or more cases of breast cancer (60). 80% of the *BRCA1* mutations result in abnormal truncation of the protein. *BRCA2* functions are similar to those of *BRCA1* and most of the mutations found in this gene also result in premature truncation of the protein. Although mutations in *BRCA1* and *BRCA2* genes are the most frequently identified causes in familial breast cancer, they are rarely found in sporadic cancer.

Additional tumor suppressor genes such as *RB*, *ATM* and *CHK2* have been also related to human breast cancer development.

2.3. Transgenic mouse models used in breast cancer research

Transgenic mice have been widely used to examine the effects of the loss of tumor suppressor genes known to have significant roles in human breast cancer, including *P53* (115), *BRCA1* (209) and *PTEN* (111); and the effects of gain of function in oncogenes such as *ERBB2* (69), *MYC*, *HRAS* (164) and *CCDN1* (214). These models have confirmed the role of these genes in breast carcinogenesis.

Most mice models use the mouse mammary tumor virus (MMTV) or the whey acidic protein (WAP) as promoters to drive expression of an oncogene because their expression is restricted to the mammary gland, being highly active in luminal epithelial cells. In addition, some groups have used a keratin 14 (K14) promoter in order to direct the expression to basal cells and study lobular breast cancers (43). However, transgene expression in these models usually extends throughout most of the ductal or lobular cells. Therefore, they are not good models of human breast cancer development; in which oncogene expression or loss of tumor suppressor function occur in a limited number of cells.

Later, new transgenic models that combined the transgene expression control by these promoters (MMTV, WAP, K14 and others) with tetracycline (tet)- regulatable transgenes, Cre recombinase-mediated somatic gene deletion or activation (192) and inducible forms of Cre recombinase (by using fusion proteins with mutated oestrogen or progesterone receptors) (168) were also used to improve the spatial and temporal control of transgene expression.

Although there are some histological differences in mammary gland tumors between human patients and animal models (31), many similarities between mouse and human have been identified: tumor formation resulted from multiple genetic mutations, tumors contained regions that resembled human breast cancer and genes associated with human cancer caused cancer in mice. Main drawbacks include that most tumors metastasize only to the lungs and that nearly all are hormone independent, while 70% of human breast cancers are hormone dependent. In addition, a comparative genomics approach to evaluate the molecular profiles of the most common transgenic mice used to study breast cancer comparing them with human tumor molecular expression profiles indicated that none of them was representative of the ER+ luminal A human subtype (74) .

2.3.1. Trp53 mouse models of mammary gland carcinogenesis

Homozygous *Trp53* knockout mice develop spontaneous tumors that are mostly lymphomas and sarcomas before they reach 6 months of age and succumb to them by 10 months of age (45), while heterozygous mice that carry one wild type *Trp53* allele rarely develop tumors before they are 9 months old. In addition, these p53+/- mice exhibit a wider array of tumors, but most of them consist of soft tissue sarcomas and lymphomas (191). Since early death due to lymphomas was supposed to be masking the development of breast tumors, several studies were done transplanting the mammary gland of p53-/- mice to a cleared fat pad of wild type p53+/+ mice (88). However the influence of the transplant process in the tumor development was not clear.

To overcome this problem several experiments were done with conditional expression of mutant forms of *Trp53* in epithelial cells of mammary glands. For example, when expressing conditionally a mutant R270H *Trp*53 allele in mammary glands using the WAP promoter, ER positive mammary gland tumors were obtained (205). A similar experiment using a mutant

Trp53 that was expressed under the MMTV or WAP promoters obtained spontaneous mammary gland tumors, in which a subset of them was ER positive and both lines MMTV and WAP transgenic mice produced metastasis to lung or liver (115).

3. THE SLEEPING BEAUTY TRANSPOSON SYSTEM

3.1. Insertional mutagenesis in cancer research

Mouse models are widely used in cancer gene discovery due to several reasons: their small size, ability to breed in captivity, a relatively short lifespan and because they display extensive physiological and molecular similarities to humans. In addition, their genome has been entirely sequenced. Genetically engineered mouse models have allowed the study of gene function in vivo and the creation of models that recapitulate human cancer. Transgenic mice (created by pronuclear injection of a transgene or by embryonic stem cells manipulation) have been used for studying the effect of overexpressing an oncogene or inactivating a tumor suppressor gene. Transgenic constructs can be designed to restrict transgene expression to specific tissues and to control when and how long the transgene is expressed by adding specific regulatory elements.

These models have been useful to study the molecular function of a particular gene or a few genes, but they present some disadvantages when applied to the study of cancer biology. First, tumors develop through a multistep process, where mutations are gradually acquired, while in these models, the genes of interest are mutated simultaneously. Second, the mutation of interest is present in every cell of a tissue, while natural tumors develop from a single mutated cell. And finally, many mouse models show high penetrance and short latency, and hence the resulting tumors may not present many of the cooperating events that natural tumors acquire with time.

To overcome these disadvantages, insertional mutagenesis has been proposed as an alternative tool for cancer gene discovery in the mouse. Forward genetic screens using somatic mutagens allows the identification of cancer candidate genes without the need of previous information about the gene or tumor type, facilitating the determination of new genes involved in carcinogenesis. Different mutagens have been used for this purpose:

Chemical mutagenesis using agents such as N-ethyl-N-nitrosourea (ENU) is a highly efficient way of inducing tumors in mice, but the causal mutations can be hard to identify.

Slow transforming retroviruses, like murine leukemia virus (MuLV) or mouse mammary tumour virus (MMTV) have been widely used for insertional mutagenesis in the mouse. Retroviruses induce tumors by inserting themselves into the mouse genome, which may lead to deregulation of a proto-oncogene or to inactivation of a tumor suppressor gene. Mutations that confer a growth advantage to the infected cell will be selected, and this may recapitulate the multistep progression of human tumors. These retroviruses contain viral genes flanked by two

long terminal repeats (LTRs) (**Figure 10**) Elements within the LTRs drive expression of the viral genes but can also disrupt the expression of host genes.

Figure 10. Viral genes are flanked by LTRs, which consist of three parts: U3, R and U5 (187) . U3 contains the enhancer and promoter sequences that drive viral transcription, the R domain includes the 5´capping sequences and the poly A signal. Gag, pro, pol and env encode the viral components required for assembly of viral particles. (Modified from Uren (187))

Retroviruses can mutate host genes in a number of different ways. The most common mechanism is enhancer mutation, where the enhancer included in the LTRs upregulates expression of host genes. An alternative mechanism of mutagenesis is promoter mutation, where the retrovirus inserts in the sense orientation into the promoter region of a host gene, producing chimeric transcripts between the retrovirus and the murine gene. A third mechanism is the production of a truncating mutation by intragenic insertion, causing premature termination of gene transcription, resulting in either gene upregulation or inactivation (**Figure 11**)

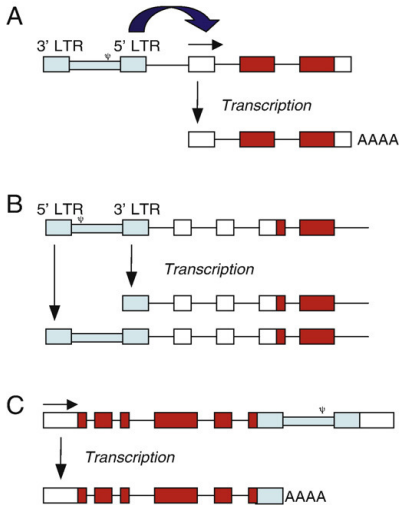

Figure 11. Retrovirus mechanisms of mutagenesis. The provirus is shown in blue and coding and non-coding exons are shown in red and white, respectively. A) Enhancer mutation. B) Promoter mutation C) Truncating mutation (Modified from Mattison (123))

According to this behavior, it is possible to identify either oncogenes or tumor suppressor genes by retroviral insertion mutagenesis. Retroviral insertion sites in the tumors mark the location of candidate mouse cancer genes, which are then identified through cloning and sequencing of the insertion sites. Hundreds of candidate genes have been identified in recent years using this approach (2). However, there are several limitations for the use of retrovirus in insertional

mutagenesis. The major limitation consists on its cellular tropism, limiting their mutagenesis to specific tissues such as mammary gland for MMTV and hematopoietic tissue for MuLV, so they are less efficient for cancer gene screening in other tumors. In addition, MuLV and other retroviruses do not efficiently infect non-dividing and slowly replicating cells and they do not infect tissues that have a basement membrane or mucin layer (197, 210). And finally, they show a strong preference for integrating near the promoter region of actively transcribed genes, which makes it difficult to efficiently mutagenize the entire genome.

Transposon-based insertional mutagenesis provides an alternative to overcome the drawbacks of retroviruses for cancer gene discovery in the mouse. Transposons are DNA mobile elements that change their position within the genome. There are two types of transposons depending on their mechanisms of mobilization.

Class I (retrotransposons) are mobilized using a RNA intermediate. They are transcribed and then reverse transcribed by enzymes encoded by the retrotransposon. The resulting DNA copies are mobilized into new sites in the genome.

Class II refers to DNA transposons. They are mobilized using a "cut and paste transposition mechanism", in which a transposase (encoded by the DNA transposon) catalyses the excision of the transposon from its original location in a given chromosome and then promotes its reintegration elsewhere in the genome. DNA transposons are also classified as autonomous, when they encode an active transposase enzyme and non-autonomous when they lack a functional transposase enzyme that needs to be supplied in *trans* for its transposition.

Although transposon mutagenesis has been successfully used in genomic screens in organisms such as *Drosophila melanogaster*, *Caenorhabditis elegans* and plants (15), (35), it has not been used in vertebrates until recently.

The *Sleeping Beauty* (SB) transposon is a synthetic transposon shown to have activity in mice (84). The SB transposon is a member of the family of Tc1/ mariner transposable elements that transposes by a "cut and paste" mechanism and was originally identified as a dormant transposable element in the genome of several species of salmonid fish. Directed mutagenesis was used to correct mutations that silenced the activity of the transposase enzyme (84), resulting in a transposon active in fish, but also in mouse and human cells (118), (85), (212), (52), (51). The SB transposon is flanked by two inverted repeats, left (IRL) and right (IRR), of ~230 bp each, flanking a cargo sequence. The cargo sequence of the transposon can be any sequence of choice. However, transposition efficiency decreases with increasing cargo sizes, being 2 kbp the optimal transposon cargo size (64). Transposition occurs by binding of the transposase enzyme to specific binding sites within the IRL/IRR elements and subsequent excision of the transposon from the donor site and insertion elsewhere in the genome at a TA dinucleotide target site. The TA site is duplicated at the insertion site and the DNA breaks generated at the donor site are also repaired by the transposase (**Figure 12**).

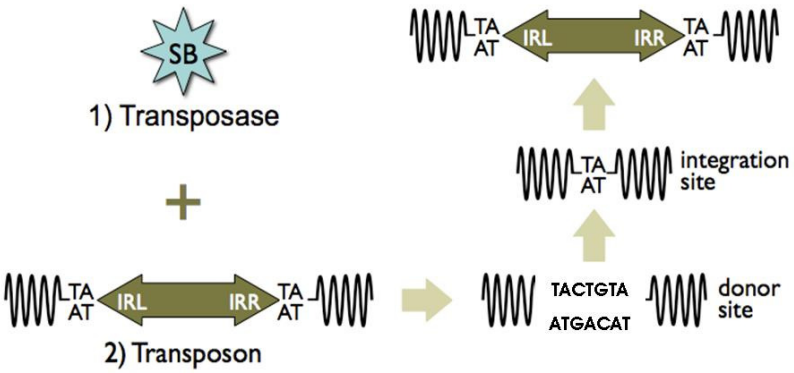

Figure 12. The Sleeping Beauty (SB) transposon system. The SB transposon system consists of two parts: the transposase enzyme SB and the transposon. When both elements are found within the same host cell nucleus, the SB can bind to the inverted repeats of the transposon (IRR and IRL), excise the transposon from its original site and mediate insertion into a new TA dinucleotide site in the genome. The TA site is duplicated upon insertion, and flanks each end of the transposon at the insertion site and remains after the DNA breaks are repaired. The repair often leaves behind a footprint, for example (TACTGTA). Modified from Dupuy (49)

3.2 The T2Onc2 transposon

The first experiments performed using the SB system in mice involved the generation of transgenic mice that expressed the transposase enzyme specifically in the mouse germ line (52). These experiments resulted in such a low rate of transposition in the germ line that it was considered not to be a practical approach in cancer gene studies in the mouse, since cellular transformation into cancer cells requires several independent gene mutations. Transposon mutagenesis in somatic cells was the following approach used for cancer gene discovery. To this aim, the T2/Onc2 transposon was developed. T2/Onc2 is a modified SB transposon that was designed to achieve higher rates of transformation in somatic cells of the mouse (50). Its cargo was designed to mimic retroviral insertion mutagenesis, being engineered with splice acceptors and polyadenylation sequences in both orientations to disrupt the expression of the genes in which they land. It also contains sequences from the 5´LTR of the murine stem cell virus (MSCV LTR) to serve as promoter/enhancer elements to drive expression of nearby genes. The MSCV LTR is also followed by a splice donor site (SD). These characteristics make this transposon an adequate insertional mutagen for cancer gene identification because it can mutate both oncogenes and tumor suppressor genes. The splice acceptors (SA) and polyadenylation sequences (pA) on both strands allow the transposon to be able to truncate tumor suppressor gene transcripts when inserted in either the forward or reverse orientation. The promoter and the splice donor site (SD) serve to initiate transcription and splice into downstream exons to drive overexpression of oncogenes (**Figure 13**)

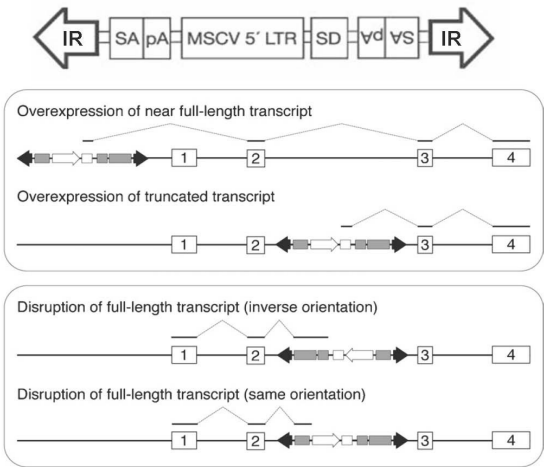

Figure 13. **T2/Onc2 transposon structure**. The T2/Onc2 transposon contains two inverted repeats (IR); the enhancer and promoter of MSCV 5' LTR, and the splice donor (SD) and acceptor sites (SA). With this structure, it can cause overexpression, disruption in the expression or expression of truncated forms of the targeted genes. Modified from Dupuy (49)

3.3. Sleeping Beauty models of cancer

Insertional mutagenesis in mice for cancer gene discovery using the SB transposon involves the use of two mouse strains, one containing the transposase transgene and another containing a multicopy concatemer of the transposon transgene. Transposon mobilization will therefore occur in the resulting offspring carrying both transgenes in their genome

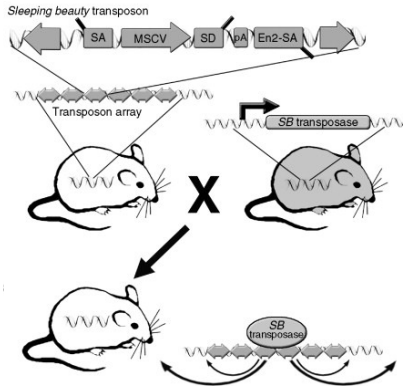

Figure 14. Transposon insertional mutagenesis in mice using the SB transposon. The progeny of transposase and transposon transgenic mice express the transposase causing mobilization of the transposon in the genome. Modified from Walrath JC (194)

Several different approaches using SB insertional mutagenesis have been done in order to find candidate cancer genes in a variety of different cancer types. Currently, three different mutagenic transposons T2/Onc, T2/Onc2 and T2/Onc3 have been used in combination with two different transposase transgenes (SB10 and SB11) to model cancer in mice. Promoter sequences in the transposon cargo and transposase expression control have been modified depending on the cancer of study. The first experiments with transposon insertional mutagenesis in mice for

cancer gene discovery involved the use of the modified transposons T2/Onc and T2/Onc2 (33), (50). Both transposons share the same structure, being the only difference that T2/Onc contains a larger fragment of the *Engrailed* 2 splice acceptor (En2). Transgenic lines were established harboring multiple copies of T2/Onc or T2/Onc2 transposons within a concatemer carrying between 20-30 and 150-300 copies respectively. These mice were mated to transgenic mice expressing the transposase gene in order to obtain mice that underwent transposition events. T2/Onc mice were bred to transgenic mice carrying the SB10 transposase transgene expressed under the control of the ubiquitous CAGGS promoter (33), while T2/Onc2 transgenic mice were crossed to transgenic mice carrying the more active SB11 transposase version (50) inserted into the Rosa26 locus (Rosa26-SB11), also ubiquitously expressed. Double transgenic mice containing the T2/Onc transposon did not present spontaneous tumor formation. However, T2/Onc was able to accelerate sarcomas tumor formation in mice deficient for the p19 *Arf* tumor suppressor. Double transgenic mice offspring containing the T2/Onc2 transposon and the SB11 transposase suffered from embryonic lethality. All the animals that survived developed aggressive lymphomas within 10 weeks of age (50). When Rosa26 SB11 mice were bred to T2/Onc mice, the offspring died at a mean age of 6 months, due primarily to lymphoma/leukaemia. Some of them also harbored infiltrating gliomas (32) and hyperproliferative prostate lesions (142). In these experiments using T2/Onc and T2/Onc2 transposons, clonal transposon insertions were found associated to tumor development and progression. In addition, at difference to what happens in retroviral insertional mutagenesis, transposon insertions were widely distributed across the genome, being the only exception to this assumption the "local hopping" phenomenon, which is defined as the tendency of the SB transposon to transpose to sites close to the concatemer. Subsequent experiments involved a conditional SB11 transposase that allowed the study of insertional mutagenesis in a broad range of tumors. A conditional Rosa 26-SB11 transposase allele was generated in which loxP sites flanked an upstream EGFP cassette (**Figure 15**). In this way, SB11 was only expressed when Cre recombinase deleted the floxed EGFP cassette (**Figure 15**). Using this approach, candidate cancer genes were identified in B-cell malignancies (53), colorectal cancer (170), hepatocellular carcinoma (95) and pancreatic adenocarcinoma (121)

Figure 15. Rosa-SB11 allele designed for condicional transposase expression. (Modified from Dupuy (53))

The recently developed T2/Onc3 transgene is identical to T2/Onc2 except that the MSCV 5'LTR was replaced by the CMV enhancer/ chicken b-actin promoter (CAG). This promoter is strongly expressed in a variety of cell types including epithelial cells. Experiments using

transgenic mice harboring 10-20 copies of this transposon that were bred to Rosa26-SB11 mice obtained a double transgenic offspring that developed a variety of malignancies, being the most common hepatocellular carcinoma liver, lung adenomas and SCC (53)

4. IDENTIFICATION OF CANDIDATE CANCER GENES

4.1. Identification of SB-tagged genes

SB transposon insertion sites within the mouse genome can be identified because the transposon sequence is different from the murine genome. Therefore, mouse sequences adjacent to the insertion sites can be amplified by ligation-mediated PCR methods (LM- PCR) using primers targeting the IRR and the IRL within the transposons (104). The PCR products are then sequenced and used to map the transposon insertion sites to the mouse reference genome, so the gene or genes likely to be affected by each transposon insertion can be identified, as well as the orientation of the transposon with respect to the gene. The first experiments performed with SB transposons as mutagenic agents followed a shotgun cloning approach where a plasmid library of the LM-PCR products from each individual tumor was created. After that, individual clones from each library were sequenced by Sanger methods to identify the transposon-genomic DNA junctions found within each tumor (50), (33), (32).

However, this technique presented notable limitations, such as laborious library production and a limited number of independent clones that could be sequenced..In order to overcome the shotgun cloning limitations, the technique was replaced with a novel method that generates LM-PCR products that can be directly sequenced on the Genome Sequence FLX platform (454/Roche) (53), (95), (170), (16). This approach represented a 3-fold increase in the sequence coverage over the traditional Sanger sequencing. In addition, recent work has shown that many hundreds of independent insertion events can be identified in each tumor sample (170); so, it is necessary to obtain enough sequence depth in order to identify all the insertion events present in the tumor samples. The pyrosequencing 454/Roche platform was chosen for transposon insertion sites identification because its longer read length (>100bp) allowed barcoding the samples, so multiple samples could be processed in a single sequence run, making at the same time possible both the verification of the transposon structure and the adequate mapping of the transposon/mouse genomic junction sequence.

However, it has been recently seen that pyrosequencing data obtained from repeated analysis of the same samples only identify part of the original insertion sites present in each tumor sample (171). Improvements in the annotation of short sequences and the sequence read length for the Illumina sequencing platform (~75bp) made transposon insertion site sequencing possible using

this technology, that in adidtion to being more economical, results in a greater sequence depth than 454/Roche pyrosequencing.

4.2. Determination of common insertion sites

Common insertion sites (CIS) are genomic regions that have been hit by transposon insertions in multiple tumors. Since insertions occur independently in each tumor, it is unlikely to find insertions in the same locus from multiple tumors by chance alone. These insertions hotspots are likely the result of selective advantage as a consequence of the presence of these mutated loci in tumor cells, so CIS regions probably represent candidate cancer genes. Two statistical methods have been developed for CIS identification: the Gaussian Kernel Convolution and the Monte Carlo simulation. Both methods calculate the expected number of transposon integrations, taking into account the exact distribution of TA dinucleotides across the genome and determine if insertion sites found in the tumors have been targeted more than expected by chance. One thing that is necessary to do before CIS analysis is the removal of local hopping transposition events (i.e., discard integrations in the same chromosome in which the concatemer is originally located), to prevent false CIS identification.

Analysis of CIS in *Sleeping Beauty* transposon-induced tumors has identified both known and novel cancer causing genes, both oncogenes and tumor suppressor genes. Frequently, the information obtained in transposon mutated mouse models has been validated by the study of human tumoral samples, thus confirming the driver nature of the characterized gene alterations in human cancer development.

Examples of important cancer genes characterized in transposon-mediated approaches are *Notch1* in SB-induced T-cell lymphoma (50), *Apc* in colorectal cancer (170), *Braf* in sarcomas (33), *Pde4D* in prostate cancer (142), *EGFR* in hepatocellular carcinoma (95) and *Csf1* in astrocytomas (16).

OBJECTIVES

Many forms of Cancer arise as a result of somatic mutations in genes regulating growth and differentiation of cells. In most tumors, multiple genes are usually affected. The number and nature of genes needed to drive cancer development can vary between different tumor types and recent cancer genome studies have determined that the number of genes with transforming potential is greater than expected. In addition, there are many passenger mutations in the tumors that are not casually implicated in oncogenesis. Therefore, it is still necessary to define more precisely which combination of mutations is required to drive tumor growth. Transposon-mediated mutagenesis allows the development of cancer and the identification of driver mutations that contribute to carcinogenesis.

The main purpose of this study was to use the Sleeping Beauty transposon system to identify novel candidate cancer genes that contribute to human skin cancer. The research objectives were:

- Generation of transgenic mice containing mutagenic transposons active in skin, other stratified epithelia and exocrine glands, both in wild type and cancer-prone genetic backgrounds, and verification of the correct activation of the system in the transgenic mice.
- Generation of tumors in these mice, and analysis of the histopathological features of the resulting tumors.
- Identification of the transposon-tagged candidate cancer genes within the mouse tumors through the use of next-generation sequencing.
- Validation of the candidate mouse cancer genes in human tumor samples

MATERIALS AND METHODS

1. GENERATION OF TRANSGENIC MICE LINES

K5-SB11 (named as SB) transgenic mice had been generated previously in our laboratory by microinjection of a K5-SB11 construct, including the SB11 transposase (a generous gift of Dr. Perry Hackett, University of Minnesota) under the control of the regulatory sequences of the bovine keratin K5 gene (143) into (C57BL/6J X DBA/2J) F2 (B6D2) embryos using standard techniques (75). Transgenic lines were established by crossing the transgenic mice with B6D2F1 mice. Heterozygous T2Onc2 mice (named as T2: lines 6070 and 6113), containing several hundred copies of T2Onc2 transposon in chromosomes 4 and 1 respectively) were provided by Dr. Adam J. Dupuy, University of Iowa(50).

Double K5-SB11/T2Onc2 (named as SB/T2) transgenic mice, containing both the SB11 and T2Onc2 transgenes were generated by interbreeding of heterozygous SB and T2 mice.

Triple K5-SB11/T2Onc2/Tg.AC (SB/T2/Tg.AC) transgenic mice were generated by mating heterozygous SB/T2 mice to Tg.AC heterozygous mice that were acquired from Taconic Farm Inc (Germantown, NY, USA). The resultant SB/T2/Tg.AC mice contained the three transgenes (SB11, T2Onc2 and Tg.AC) in a 75% B6D2 / 25% FVB genetic background.

Heterozygous p53+/- transgenic mice were obtained by mating conditional mutants *Trp53*$^{F2\text{-}10}$ mice, (that carried floxed *Trp53* alleles in a FVB genetic background provided by the laboratory of Dr. Anton Berns (Netherlands Cancer Institute),(91)) to K5-Cre transgenic female mice (that cause ubiquitous cre-mediated recombination of floxed alleles in preimplantation embryos) (144). Triple transgenic K5-SB11/T2Onc2/p53+/- mice (SB/T2/p53+/-) were generated by mating heterozygous SB/T2 to heterozygous p53+/- mice.

In these crosses, different genetic combinations lacking the transgenes T2Onc2 (SB/Tg.AC and SB/p53+/-), the transposase SB11 (T2/Tg.AC and T2/p53+/-) or both SB and T2Onc2 trasgenes (p53 +/-) were also generated as control mice for the experiments.

Mice were maintained on standard rodent diet and kept under a 12h light-12h dark cycle. All experimental procedures were performed according to European and Spanish laws and regulations (European Convention ETS 123 on the use and protection of the vertebrate mammals used in experimentation and other scientific purposes; Spanish R.D 1201/2005 of the Ministry of Agricultural, Food and Fisheries on the protection and use of animals in scientific research) and approved by our institution Ethics Committee. Mice were euthanized in a CO2 chamber for tumor extraction or when they presented notable deterioration, following FELASA and CIEMAT regulations.

2. ISOLATION OF GENOMIC DNA

2.1. Tail tips biopsies of mice

Tail tip biopsies were digested in 700µL of lysis buffer (100mM Tris-HCl pH 8.5, 5mM EDTA, 0.2% SDS, 200mM NaCl and 100µg proteinase K/ml) overnight at 55ºC. DNA was precipitated with 1 volume of isopropanol, washed with 70% ethanol and resuspended in TE buffer (10mM Tris-Cl, 1mM EDTA, pH: 8.0)

2.2. Epidermal and dermal DNAs

In order to isolate the epidermal and dermal DNAs from newborn or adult animals, back skin or tail skin, respectively were incubated for 2 min at 60ºC and epidermises were then mechanically separated from dermal tissue under a dissecting microscope. After that, dermal and epidermal DNAs were extracted as described for tail biopsies.

2.3. Tumors

Tumor samples were removed from the euthanized animals, minced and digested in 700 µL of lysis buffer (50 mM Tris-HCl pH 8,5, 100 mM EDTA, 1% SDS, 100 mM NaCl and 100 µg/ml proteinase K) overnight at 55ºC. DNA was then precipitated with 250 µL of saturated NaCl and 500 µL of isopropanol, and washed with 70% ethanol. The DNA was resuspended in TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0).

For some tumor samples, the QIAGEN Genomic DNA 20/G kit isolation was used, according to the manufacturer protocol.

3. GENOTYPING OF TRANSGENIC MICE

Mice were genotyped by PCR analysis of tail DNA. The primers used for DNA amplification are listed below (**Figure 17**). For amplification of the SB11 and T2Onc2 transgenes, the PCR program used consisted of an initial denaturing step of 94ºC for 3 min; 40 cycles of denaturing at 94ºC for 15s, annealing at 55ºC for 30s and extension at 72ºC for 45s; followed by a final extension at 72ºC for 3 min.

For PCR amplification of the Tg.AC transgene and the *p53* gene,, conditions consisted of an initial denaturing step of 94ºC for 5 min; 29 cycles of denaturing at 94ºC for 30s, annealing at 58ºC for 30s and extension at 72ºC for 50s; followed by a final extension at 72ºC for 5min.

PCR products were then separated on a 1.7 % agarose electrophoresis gel in TBE 1X (Tris-borate-EDTA: 89mM Tris-borate, 2mM EDTA, pH 8) and 1µg/mL ethidium bromide (Sigma). A DNA ladder was added to estimate the size of the amplified PCR products (DNA Molecular

Weight Marker IX, Roche and Lambda DNA marker, Bioron). The DNA was visualized using a UV transilluminator and Quantity One software (Biorad)

3.1. Genotyping PCR reaction

Forward and reverse primers to detect SB11 and T2Onc2 transgenes were combined in an equal mix to facilitate detection of both transgenes in only one PCR reaction. This was possible because the PCR amplified fragments for both genes are of different size. PCR reagents and *Taq* polymerase were purchased from Invitrogen.

Reaction	**SB11/T2Onc2**	**p53**
DNA	1 µL	1 µL
Water	12.8 µL	11.8 µL
Buffer 10x	2 µL	2 µL
Mg 50mM	1 µL	1 µL
Reverse primer	0.5 µL	1 µL
Forward primer	0.5 µL	1 µL
dNTPs 2mM	2 µL	2 µL
Taq polymerase	0.2 µL	0.2 µL

Figure 16. PCR reaction setup for genotyping

GENES	PRIMERS	SEQUENCE (5'-3')
SB11	b-globin 1	TTCAGGGTGTTGTTTAGAATGG
	b-globin 2	CAATAAGAATATTTCCACGCCA
T2ONC2	T2Onc2 Left	CAGTTGAAGTCGGAAGTTTA
	T2Onc2 Right	GGAATTGTGATACAGTGAAT
Tg.AC	Forward	TGGCATTTCTTCTGAGCAA
	Reverse	TTGGACAAACTACCTACAGAGAT
p53 null	10R	GAAGACAGAAAAGGGGAGGG
	1F	CACAAAAACAGGTTAAACCCAG

Figure 17. Primers used for genotyping

4. EXCISION PCR

Excision PCR was used to study transposon mobilization in SB/T2 double transgenic animals. The primers used were Excision1 (5´- TGTGCTGCAAGGCGATTA -3') and Excision2 (5´-ACCATGATTACGCCAAGC -3´). Conditions for PCR were: initial denaturing step of 95ºC for 5 min; 40 cycles of denaturing at 95ºC for 30s; annealing at 50ºC for 30s and extension at 72ºC for 1min20s; followed by a final extension at 72ºC for 3 min. PCR products were separated on a 1% agarose electrophoresis gel following the same procedure described in 3.

5. RNA ISOLATION

5.1 Mouse tissue

Mouse normal skin or tumor samples isolated from transgenic mice were disrupted and homogenized in 1.5 mL Trizol reagent (Invitrogen) using a Polytron homogenizer. Then, for phase separation, 0.3 mL of CH_3Cl was added to samples. After that, samples were vortexed vigorously for 15s and incubated at 5-10 min room temperature. Next, samples were centrifuged (14000 rpm for 10 min at 4ºC) and the aqueous phase was preserved for total RNA isolation. RNA was precipitated with isopropanol and washed with 70% ethanol. RNA was resuspended in 100µL of RNase- free water.

5.2. Human tissue

Fresh tissue samples were obtained from ambulatory patients in the surgical service of the Hospital Gregorio Marañón. All samples were obtained with an informed consent, according to the ethical guidelines. A 4mm punch biopsy was obtained from the center of the tumor tissue, and another biopsy was taken from healthy skin outside the tumor margin. The margins of the tumors were sent to the pathology service to confirm that there was not malignant tissue. Biopsies were immersed in RNAlater (Ambion) for preservation immediately after extraction. After 24h, RNA later was removed and samples were frozen at -70ºC until processed. Frozen samples were then disrupted and homogenized using Trizol reagent (Invitrogen) using a Polytron homogenizer, as described above (5.1). Total RNA was then purified using RNAeasy columns (QIAGEN) and RNA quality was checked using phosphate agarose gels, spectophotometry analysis and an Agilent Bioanalyzer 2100.

6. NORTHERN BLOT

20 µg of RNA isolated from mice epidermis were run on agarose gels using MOPS buffer (SIGMA) and 6% formaldehyde. Then, the gel was transferred overnight to nylon membranes (Amersham, GE Healthcare) using 10x SSC (saline-sodium citrate) buffer. After that, the membrane was UV-irradiated and kept for 2 hours at 70 ºC. The membrane was then hybridized with a SB11 DNA probe and a 7S RNA probe to normalize loading differences.

7. SOUTHERN BLOT

For the analysis of transposition in tissues and tumors from SB/T2 mice, genomic DNA was digested with *Bam*H1 and blotted using standard techniques(167). The membrane was hybridized with a 278bp PCR-amplified-probe from the region of T2/Onc2, between the IRL and the MSCV promoter. Hybridation with *Thy1* was performed to normalize loading differences

8. REVERSE TRANSCRIPTION

The cDNAs were prepared from 2 µg of total RNA isolated from human tumors using the High Capacity RT kit (Applied Biosystems) according to the manufarturer's instructions.

9. REAL TIME-PCR

Quantitative real-time PCR was performed using an ABI 7500 Fast Instrument (Applied Biosystems) and predesigned probes from Taqman Gene Expression assays (Applied Biosciences) or from the Universal Probe Library (Roche). Probe selection and primer design was done using the Probe Finder Assay software (Roche) for every gene in the expression analysis study. Normalization was performed using the geometrical means of four genes (*TBP*, *βACT, HPRT* and *GUSB*) according to the recommended procedure (189). All samples were assayed at least twice and statistical analysis was performed using SPSS and MeV (153).

	PRIMERS (5'-3')		PROBES (5'-3')
GENE	FORWARD	REVERSE	
WAC	CCTCCAACACCGACTACCC	GTTTCTTCAGTTCAGTGGAATG	# 66
TNRC6B	GGTCCACTGTTTAAAGCGAGA	GGCCAATACCTGGATATGGTT	# 12
PARD3	CATTCAGGATGGCCGACT	TGGGATTTGCCCACTAAATCT	# 51
NSD1	GTCTGGGTAAAAGTTGGACGA	AGGAACAGCTCGAGGATGG	# 49
ADAM10	ATATTACGGAACACGAGAAGCTG	TCAATCGCTTTAACATGACTGG	# 61
TRPS1	CGTGTGGCCTCTACCAGAAG	GGGTTAAGGGCGTTTCTTG	# 63
NOTCH1	CGCACAAGGTGTCTTCCAG	AGGATCAGTGGCGTCGTG	# 85
FAT1	CCTTCTGACAGCGACTCCAT	CAGGGATCAAGATCCACCAC	# 33
ATPA2A2	AACGTCGGGGAAGTTGTCT	GAATCAAAGCCTCGGGAAAT	# 66
SYNCRYP	CACTGGACAGAGGAAGTATGG	GATCTTTCCCACAAATATCTCA	# 10
TP63	CCTCCAACACCGACTACCC	GTTTCTTCAGTTCAGTGGAATACG	# 11
βACT	ATTGGCAATGAGCGGTTCC	CACAGGACTCCATGCCCA	CCCTGAGGCACTCTTCCAGCCTTCC

GENE	PRIMERS AND PROBE ABI REFERENCE NUMBER
GLI2	Hs 0025797_m1
TBP	4333769-0910013
HPRT	4333768-0911024
GUSB	4333767-0812017

Figure 18. Primers and probes from Universal Probes Library and ABI used for Real Time

10. WESTERN BLOT

Protein extracts from mouse tissues were obtained by homogenization in a liquid N2-frozen mortar followed by 3 freezing/unfreezing cycles in lysis buffer (Hepes 200 mM pH7.9, glycerol 25%, NaCl 0.4M, EDTA 1mM, EGTA 1mM and Nonidet P40 1%) supplemented with proteases inhibitors: DTT (2.5mM) aprotinin/leupeptin (10 mg/µL), PMSF (1mM), orthovanadate (1mM) and fosfatases inhibitors: NaPPi (1mM) and FNa (5 mM) . Protein concentration was determined using the colorimetric Bradford method (BioRad). Denatured and reduced sample proteins were separated by 4-12% gradient polyacrilamyde gel electrophoresis (SDS-PAGE) using a NuPAGE MOPS running buffer (Invitrogen). 40µg of the total protein extract from each sample and a protein ladder (BioRad) were used for gel electrophoresis. Samples were then transferred to a nitrocellulose membrane (Invitrogen) using transfer buffer (33 mM Tris, 19,4mM Glycine) with 10% of methanol. Membranes were incubated with blocking buffer (5% milk and 0.1% Tween 20 in TBS (Tris-Cl 20 mM, 137 mM NaCl, pH 7.6) for 1h at room temperature. A monoclonal anti-SB antibody (R&D Systems) was used at 1:500 dilution and incubated o/n at 4ºC with shaking. Peroxidase-conjugated AffiniPure Donkey Anti-Mouse (Jackson) was used as a secondary antibody at 1:2000 dilution. β-Actin antibody (Santa Cruz) was used as a loading control at 1:2000 dilution. These antibodies were incubated 1h at room temperature with shaking. All primary and secondary antibodies were diluted in a solution of 0.1% Tween 20 and 1% of BSA in TBS (Tris-Cl 20 mM, 137 mM NaCl, pH 7.6). Nitrocellulose membranes were washed 3 times with TBS-Tween 1% between antibody assays. The Super Signal Wet Pico Chemiluminescent Substrate (ThermoScientific) was used for secondary antibody detection, following the manufacturer's instructions and the membranes were exposed on a photographic film (Amersham Hyperfilm ECL).

11. IMMUNOHISTOCHEMISTRY ANALYSIS

Isolated tissues from mouse origin were fixed in 4% formaldehyde in PBS or 70% ethanol and then paraffin-embedded. Deparaffinized sections were stained with hematoxylin and eosin. For samples fixed in formaldehyde, a 10mM citrate buffer or a EDTA buffer: 10 mM Tris Base, 1 mM EDTA Solution, 0.05% Tween 20, pH 9.0 (for NSD1) was used for antigen retrieval. Samples were the incubated in a 0.3% H_2O_2 solution to block the endogenous peroxidase activity. In addition, in order to avoid non-specific reactions, selected sections were incubated with a solution of 2.5% fetal bovine serum in PBS during 30 min at room temperature. Selected sections were then immunostained with primary antibodies (**Figure 19**) in PBS and BSA 1% at 4ºC o/n. After that, sections were incubated with the adequate secondary antibodies (**Figure 20**)

in PBS and BSA 1% for 1h at room temperature. The immunoreaction was detected using the ABC avidin-biotin-peroxydase conjugate system (ABC kit, Vectastain). DAB was used as a chromogen (Peroxidase Substrate Kit DAB, Vector Laboratories). Samples were then counterstained with hematoxylin and visualized under an optical light microscope.

12. IMMUNOFLUORESCENCE ANALYSIS

Deparaffinized sections that were fixed in ethanol were used for immunofluorescence analysis, so antigen retrieval was not necessary. Double staining was done incubating primary antibodies against CD34 and SB in the same tissue section. Skin sections were incubated with primary antibodies against CD34 and SB diluted in 4% BSA in PBS at 4ºC o/n. Biotinilated anti-rat and texas red conjugated anti-mouse secondary antibodies were diluted in 4% BSA in PBS and added to the sections, then incubated for one hour at room temperature. After three washes, anti-streptavidin FITC-conjugated secondary antibody was added to the sections and incubated for 1h at room temperature. After that, sections were mounted using a solution of Mowiol containing DAPI. Slides were seen under a Zeiss Axioplan2 imaging fluorescence microscope. Images were taken using a digital camera AxioCam MRm (Zeiss) and processed using AxioVision Rel 4.6 software and Corel Photo Paint11.

PRIMARY ANTIBODIES

ANTIBODY	SOURCE	SUPPLIER	DILUTION	REFERENCE
K5	Rabbit polyclonal	Covance	1:1000 IHC	PRB-160P
SB11	Mouse monoclonal	R&D Systems	1:50 IHC/IF	MAB2798
Nsd1	Rabbit polyclonal	Bethyl	1:50 IHC	IHC-0027
Pard3	Rabbit polyclonal	Abcam	1:50 IHC	ab64646
Ras GAP 120	Mouse monoclonal	Santa Cruz	1:50 IHC	SC-63
Neurofibromin (Nf1)	Rabbit polyclonal	Santa Cruz Biotechnologies	1:50 IHC	SC-67
CD34	Rat monoclonal	eBiosciences	1:50 IF	14-0341

Figure 19. List of primary antibodies used in Western Blot and immunohistochemistry assays

SECONDARY ANTIBODIES

TARGET SPECIES	CONJUGATE	SUPPLIER	DILUTION
Mouse	Biotin	Jackson	1:1000
Mouse	Texas Red	Jackson	1:200
Rabbit	Biotin	Jackson	1:1000
Rat	Biotin	Jackson	1:500
Streptavidin	FITC	BDPharmigen	1:250

Figure 20. List of secondary antibodies used in Western Blot and immunohistochemistry assays

13. DMBA/TPA CARCINOGENESIS ASSAY

The dorsal skin of SB/T2 and SB/T2/Tg.AC mice was shaved (approximately an area of 8 cm^2) 2 days prior the treatment. SB/T2/Tg.AC mice were treated with 12-O-tetradecanoylphorbol-13-acetate (TPA, Sigma, 5 μg in 200 μL acetone) twice a week for 20 weeks. SB/T2 mice were treated one week before promotion with TPA with a single dose of 200 nmol of 7,12-dimethylbenz(a)anthracene (DMBA, Sigma) (100 μg in 200 μL acetone). Mice were examined weekly for the presence of tumors.

14. MICE TUMOR COLLECTION

Isolated mice tumors were collected in:

- Formaldehyde (4%) in PBS or in ethanol (70%) and subsequent paraffin embedding for immunohistochemistry analysis.
- Liquid nitrogen and storage at -80 C for RNA/protein assays

15. AMPLIFICATION AND SHOTGUN CLONING OF TRANSPOSON INSERTION SITES

15.1. SplinkTA-PCR

First, 3μg of genomic DNA isolated from tumors was digested with HindIII, XhoI and PvuII (New England Biolabs, NEB) restriction endonucleases for 2h at 37ºC.

Reaction setup	
HindIII	3 μL
XhoI	3 μL
PvuII	3 μL
Buffer NEB	10 μL
BSA	1 μL
Water	Up to 100 μL
DNA	3 μg

The digestion products were purified using the QIAquick PCR purification kit (QIAGEN) according to the manufacturer's instructions. To achieve efficient ligation of genomic DNA fragments to splinkerettes, DNA fragments were modified using *Taq* polymerase to form sticky ends through the terminal non-templated adenosine addition reaction. For these adenosine addition reactions, each purified, digested DNA sample (50 µL) was incubated with the following PCR mix during 5 min for 72ºC. PCR reagents and Taq polymerase were purchased from Invitrogen.

Reaction setup	
Buffer 10x	6 µL
Mg 50mM	1,8 µL
dATPs 10mM	1,2 µL
Taq polymerase	1 µL
Digested DNA	50 µL

After this reaction, another round of DNA purification was done using the QIAquick PCR purification kit (QIAGEN).

The SplinkTA tail was generated by mixing equal molar amounts of the olignonuleotides splinkerette (the hairpin oligonucleotide) and PrimerLongTA (including an extra T) and incubating them for 3 min at 94 -100 ºC and then slowly cool down to room temperature (25ºC). After that, the SplinkTA was ligated to the extension products using T4 DNA ligase (New England Biolabs) by overnight incubation at 16ºC, followed by DNA purification with the QIAquick PCR purification kit.

Reaction setup	
Buffer	2 µL
T4 DNA ligase	2 µL
DNA	5 µL
Splinker	5 µL
Water	6 µL

Then, the resulting products were digested using *BamH1*, followed by another round of DNA purification using the QIAquick purification DNA kit (QIAGEN). This digestion prevents the amplification of transposons within the concatemer. All the reagents were from New England Biolabs.

Reaction setup	
Buffer NEB	5 µL
BSA	0,5 µL
BamH1	1 µL
Water	Up to 50 µL
DNA	3 µγ

After that, two primary PCR reactions were performed using 1µL of resulting DNA from the previous reaction in order to amplify transposon-genomic junctions from the right and left sides of the transposon. One was done using primers that annealed to the splinkerette (primerette1, complementary to PrimerLongTA) and the left inverted repeat side (IRDRL1) of the transposon and the other one included primers that annealed to the splinkerette (primerette1) and the right inverted repeat side of the transposon (IRDRR1). *Taq* polymerase and the rest of reagents were from Invitrogen.

PCR conditions were: denaturation at 95º C for 3min, followed by 10 cycles of 95º C for 15s; 70º C for 2 min; 20 more cycles of 95º C for 15s; 65 ºC for 2 min; and a final extension at 72 ºC for 7 min.

Reaction setup	
DNA	2 µL
Buffer 10x	5 µL
dNTPs 10mM	2 µL
Primerette1	0,5 µL
IRDRR1 or IRDRL1 10 µM	0,5 µL
Taq Platinum polymerase	1 µL
Mg 50mM	1,5 µL
Water	38,5 µL

Then, for each case a nested PCR was done using 2µL of DNA from the previous reactions and primers that annealed to the splinkerette (primerette 2, complementary to PrimerLongTA) and either to the transposon left side (IRDRL2) or right side (IRDRR2).

Reaction setup	
DNA	2 µL
Buffer 10x	5 µL
dNTPs 10mM	2 µL
Primerette2	0,5 µL
IRDRR2 or IRDRL2 10µM	0,5 µL
Taq Platinum polymerase	1 µL
Mg 50mM	1,5 µL
Water	38,5 µL

PRIMERS	SEQUENCE (5´-3)
Splinkerette	CATGGTTGTTAGGACTGGAGGGGAATCAATCCCCT
PrimerLongTA	CCTCCACTACGACTCACTGAGGGCAAGCAGTCCTAACAACCATGT
IRDRR1	GCTTGTGGAAGGCTACTCGAAATGTTTGACCC
IRDRL1	CTGGAATTTTCCAAGCTGTTTAAAGGCACAGTCAA
Primerette1	CCTCCACTACGACTCACTGAAGGGC
Primerette2	GGGCAAGCAGTCCTAACAACCATG
IRDRR2	CCTCCACTACGACTCACTGAAGGGC
IRDRL2	GACTTGTGTCATGCACAAAGTAGATGTCC

Figure 21. List of primers used for SplinkTA generation and PCR reactions.

15.2. Cloning and sequencing

PCR products were run in a 1.3% agarose gel and individual chosen bands were purified using the Illustra GFX PCR DNA and Gel Band purification kit (GE Healthcare) according to the manufacturer's instructions.

After that, the purified PCR products were cloned into a TOPO TA vector (pCR2.1) using the TOPO TA Cloning kit (Invitrogen), followed by transformation of XL-Blue competent cells. Colonies were grown on agar plates and positive colonies containing the DNA fragments were picked and cultured in liquid medium. Isolated DNA from cultures was then sequenced using universal primers (M13) on a 3730 DNA Analyzer (Applied Biosystems).

15.3 Sequence analysis

The resulting sequences were analyzed using the Insertional Mutagenesis Mapping and Analysis Tool (iMapper) (http://www.sanger.ac.uk/cgi-bin/teams/team113/imapper.cgi.).

Using this tool sequences were scanned and trimmed to remove the splinkerette (linker) sequence and sequences derived from the transposon insertions. Contaminating sequences derived from chimeric genomic fragments, vector or transposon concatemer were also removed. After that, resulting sequences were mapped against the mouse Ensemble genome (NCBI mm37) using a SSAHA algorithm, and these regions were then annotated. Genes were defined including 10kb of both 5' and 3' flanking regions. (**Figure 22**)

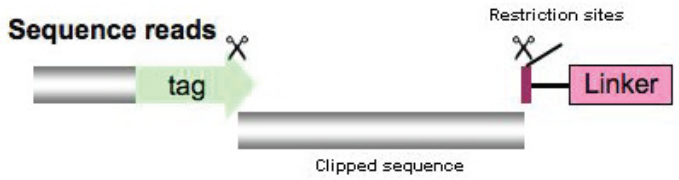

Figure 22. The iMapper tool identifies a transposon tag sequence and the restriction enzyme site within a DNA sequence read and the clipped sequenced is then mapped onto the Ensemble genome. (Modified from DJ Adams (101))

16. HIGH TROUGHPUT SEQUENCING

16.1. Pyrosequencing using 454/Roche

First, 2µg of isolated DNA from each tumour sample was digested with *Bfa*I or *Nla*III restriction enzymes to split the left (IRL) and right (IRR) inverted repeats of the transposon. The digestion was incubated o/n at 37ºC.

Reaction setup	
NlaIII or BfaI	1 µL
Buffer NEB 10x	4 µL
BSA 10x	4 µL
Water	Up to 40 µL
DNA	2 µg

These genomic-transposon junction fragments were amplified by ligation-mediated PCR (LM-PCR) using adaptors that attach to the transposon-genomic DNA junctions after digestion and facilitate the PCR amplification of these fragments.

A couple of adaptors were generated to join to each side of the transposon, one to the right side (IRR) and the other to the left side (IRL). The adaptors were generated by mixing equal molar amounts of linker+ and linker- primers and heating them to 95ºC for 5min. Then, the mixture was allowed to slowly cool down to room temperature (25ºC). After that, the adaptors were ligated to the genomic fragments using T4 DNA ligase. The reaction was performed o/n at 16 ºC.

Reaction setup	
Buffer NEB 10x	2 µL
T4 DNA ligase	1 µL
DNA	10 µL
Adaptor	1,5 µL
ATP 10mM	2 µL
Water	3,5 µL

Then, the resulting DNA was digested using *Bam*HI o/n at 37ºC, in order to prevent transposons within the concatemer from being amplified.

After that, two primary PCR reactions were performed using 2µL of resulting DNA from the ligation reaction in order to amplify transposon genomic junctions from the right and left sides of the transposon. One was done using primers that annealed to the adaptor (linker primer long) and the left inverted repeat side (IRL1) of the transposon and the other one included primers that annealed to the adaptor (linker primer long) and the right inverted repeat side of the transposon (IRR1). The *Taq* polymerase and reagents were purchased from New England Biolabs.

Reaction setup	
DNA	2 µL
Buffer 10x	5 µL
dNTPs 10mM	1 µL
Linker primer long 10µM	0,5 µL
IRR1 or IRL1 10µM	0,5 µL
Taq Platinum polymerase	0,5 µL
Mg 50mM	2 µL
Water	38,5 µL

After that, two nested PCRs were done using 2µL of a 1:50 dilution of the former PCR products, primers that annealed to the adaptor (linker pyro), and primers that contained unique sequences (A/B) which were added to the 5' ends of the linker pyro primer (A) and the transposon specific primer (B). These A/B tags were used for amplicon sequencing on the high throughput sequencer Roche/454. In addition, the transposon specific primers also contained a

10bp barcode sequence added to the 5' end of these primers and were used for tumor ID sample tracking (**Figure 23**)

Reaction setup	
Diluted primary PCR DNA product (1:50)	2 µL
Buffer 10x	10 µL
dNTPs 10mM	2 µL
IRR or IRL Barcoded primers (includes A tag) (10µM)	1 µL
Linker pyro (includes B tag) (10µM)	1 µL
Taq Platinum polymerase	1 µL
Mg 50mM	4 µL
Water	79 µL

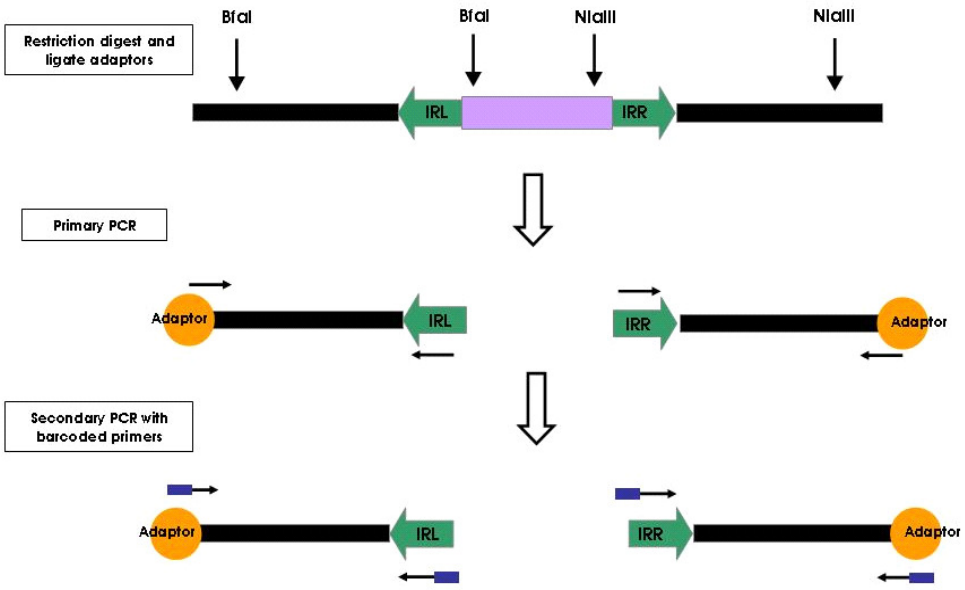

Figure 23. Scheme representing the LM PCR protocol to amplify transposon-genomic junctions from genomic DNA isolated from SB/T2 tumors. DNA is digested using *Bfa*I or *Nla*III enzymes. Double stranded adaptors are then ligated to the ends of genomic fragments. Then, a primary and secondary PCR are performed using specific barcoded primers that include the necessary sequences for direct sequencing on 454/Roche platform

PRIMERS	SEQUENCE (5´-3)
IRR linker +	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATG
IRR linker -	Phos-GTCCCTTAAGCGGAC-C3-spacer
IRL linker +	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC
IRL linker -	Phos-TAGTCCCTTAAGCGGAG-C3spacer
IRR1	GCTTGTGGAAGGCTACTCGAAATGTTTGACCC
IRL1	CTGGAATTTTCCAAGCTGTTTAAAGGCACAGTCAAC
Linker primer long	GTAATACGACTCACTATAGGGC
Linker pyro	GCTTGCCAGCCCGCTCAGAGGGCTCCGCTTAAGGGAC

Figure 24. List of primers used in the adaptor and PCR reactions. Barcoded primers are listed in supplementary materials

After that, an aliquot of 50 µL of every PCR product was analyzed on a 1.5% agarose gel using TAE; 1 mM EDTA. Later, the remaining PCR product was purified to remove excess of primers and dNTPs and the concentration of every purified PCR product was measured using a fluorometer. The PCR products were then combined in two different single tubes, each one containing either the IRL products or the IRR products. The PCR products were diluted in TE such that the final concentration of each PCR product was 200.000 molecules/mL and, shaked for 3-5 hours. After that, a final dilution that accounted for sample number (dilution of 1X for each unique barcode in the combined sample, IRR or IRL) was performed before submitting the samples for sequencing to the core facility.

16.2. lllumina sequencing

Isolated genomic DNA from tumors was quality-tested and quantified. After that, 2 µg of tumor genomic DNA was digested with *Nla*III or *Alu*I restriction enzymes. Then, IRR and IRL transposon-genomic junctions were ligated to adaptors that allowed the amplification of these fragments by PCR and digestion with *BamH*I-*H*F was performed to prevent amplification of transposons within the concatemer. These reactions, including the adaptor preparation, were carried out as described in the previous 16.1 section. The only difference in this process is that NlaIII and AluI make two cuts within the transposon and the genomic sequence, so four sets of samples were generated: *NlaIII*-IRL, *NlaII*I-IRR, *AluI*-IRR and *AluI*-IRL (**Figure 26**) Therefore, four primary PCR reactions were then performed using 2 µL of resulting DNA from the ligation reaction in order to amplify tranposon genomic junctions from right and left sides of the transposon. One was done using primers that annealed to the adaptor (linker primer long) and the left inverted repeat side (IRL) of the transposon and the other one included primers that annealed to the adaptor (linker primer long) and the right inverted repeat side of the transposon (IRR). The *Taq* polymerase and reagents were purchased from New England Biolabs.

Reaction setup	
DNA	2 µL
HF Buffer 5x (includes Mg)	5 µL
dNTPs 10mM	1 µL
Linker primer long (10µM)	2,50 µL
IRR or IRL primer (10 µM)	2,50 µL
Phusion polymerase	0,5 µL
Water	31,50 µL

Next, a secondary nested PCR was done using a primer that linked to the adaptors (linker-A2) and specific primers that included a 6bp-barcode for each sample that was used for sample tracking.

Reaction setup	
Diluted primary PCR DNA product (1:50)	2 µL
HF Buffer 5x (includes Mg)	20 µL
dNTPs 10mM	2 µL
Linker A2 primer (10µM)	2,50 µL
Barcoded IR-A1 primer	2,50 µL
Phusion polymerase	0,5 µL
Water	63 µL

After that, 25 µL of each PCR product was run on a 1.5% agarose gel using TAE buffer and products were then purified to remove excess primers and dNTPs using the QIAGEN Minielute 96 UF PCR purification kit (QIAGEN). Then, the concentration of purified products was measured using a Nanodrop spectrophotometer (ThermoScientific) and 25ng of each PCR product was pippetted into a sample pool with a final concentration of 25ng/µL. This final pool was then run on a single lane of a flow cell line on an Illumina Genome Analyzer IIx.

PRIMERS	SEQUENCE (5´-3)
IRR linker +	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATG
IRR linker -	Phos-GTCCCTTAAGCGGAC-C3-spacer
IRL linker +	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC
IRL linker -	Phos-TAGTCCCTTAAGCGGAG-C3spacer
IRR	GGATTAAATGTCAGGAATTGTGAAA
IRL	AAATTTGTGGAGTAGTTGAAAAACGA
Linker primer long	GTAATACGACTCACTATAGGGC
Linker-A2	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTAGGGCTCCGCTTAAGGGAC

Figure 25. List of primers used in the adaptor and PCR reactions

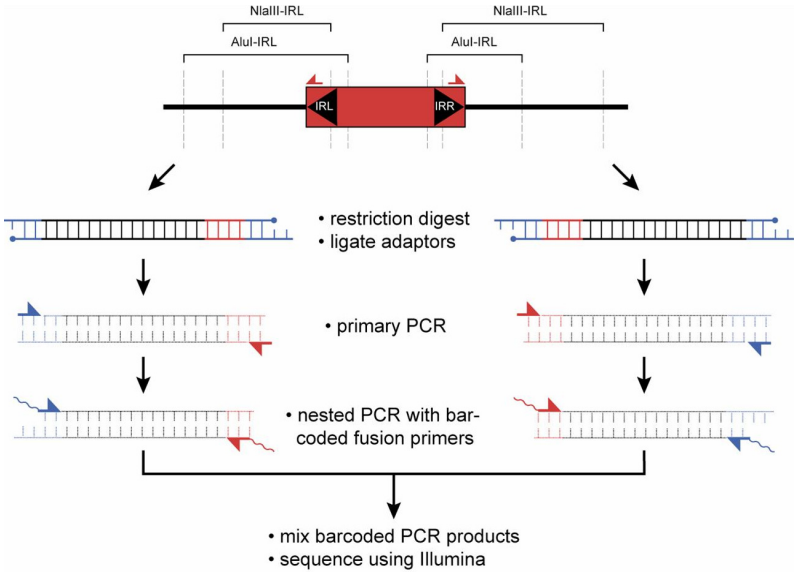

Figure 26. Scheme representing LM PCR protocol to amplify transposon-genomic junctions from tumors induced by SB mutagenesis. Genomic DNA from every tumor is digested using *AluI* or *NlaIII* enzymes. Double stranded DNA adaptors are then ligated to the end of the genomic fragments. Then, a primary and secondary PCR (nested PCR) are done using specific barcoded primers that include the necessary sequences for directed sequencing on the Illumina GAIIx platform. Modified from Ref Dupuy (53)

17. CIS IDENTIFICATION

17.1. Sequence reads analysis obtained from mammary gland tumors

The sequencing output data was analyzed using the Integration Analysis System (IAS) developed by the University of Iowa Bioinformatics Department. For this analysis, two different files were used. The first one consisted on a FASTA file containing all the sequence reads and the second one was a barcode file containing the barcode sequence information of each tumor, the tumor ID (identifier) and the information that allows identifying which sequences were derived from the left inverted repeat (IRL) or from the right inverted repeat (IRR). This barcode information was used to create individual tumor files including the information of the corresponding reads generated by sequencing. This process generated 4 independent files per tumor corresponding to the LM-PCR genomic fragments: NlaIII-IRL, NlaIII-IRR, AluI-IRR and Alu I-IRL.

After that, barcode and adaptor sequences information were trimmed from every tumor sample, in order to be ready for mapping into the mouse genome. Resultant genomic sequences were then mapped into the reference mouse genome (NCBI37/mm9) using the Bowtie algorithm (103). The mapped genome sequences were then annotated and saved into a different annotation file for every tumor that contained the tumor ID, the gene name, the gene region hit (intron or exon), the predicted effect of the transposon insertion on gene expression (disruption or activation), the distance of the insertion from the gene, the location within the chromosome, the genomic address, the % of reads derived from the insertion site and the transposon orientation relative to the gene (same or opposite).

Next, all transposon events that mapped to the donor chromosomes (chr1 from transgenic line 6113 and chr 4 from transgenic line 6070) were removed from the data, to prevent any bias related to "local hopping" phenomena. Sequences that matched to the *En2* or *Foxf2* genes were also removed from the data, since sequences of these genes are contained within the transposon structure. Then, a clonality cutoff was calculated for every annotation file in order to remove non significant background insertion sites using statistical methods (27). This process is used to identify statistically significant clonal insertion sites that are consequence of the clonal nature of tumor cells, meaning that transposon insertions that caused driving mutations will be expected to be present in most of tumor cells. Finally, these clonal insertion files identified from the four annotation files per sample generated after LM-PCR were combined in a new file for every tumor sample that contained a list of non-redundant insertion sites per tumor.

17.2. Monte Carlo simulation

A modified Monte Carlo simulation was generated containing a set of randomly selected TA sites in the mouse genome for each transposon integration that was identified. Next, 1 x 10^6 iterations of the simulation were performed. The output was used to define a common insertion site (CIS) region as the minimum genomic region in which *N* different insertions are observed to be significant when compared with the simulation ($P = 0.0001$). The CIS interval size was determined for CISs that contained between 3 and 10 insertions, and for which a single tumor contributes no more than 2 of the events.

17.3. gCIS analysis

The expected number of transposon insertion events was calculated based on the number of tumors analyzed, the number of insertion events within each tumor and the number of SB transposon target sites (TA dinucleotides) within each RefSeq transcription unit, including 10kb of promoter sequence. This analysis was performed using a chi-squared test that determined whether any RefSeq gene harbored more insertions than expected by chance.

The statistic test was used to determine the p-value, with a single degree of freedom. After that, using the Bonferroni method, a statistical threshold of $p = 2.63 \times 10^{-6}$ correction was applied to correct for multiple hypothesis testing. Any remaining RefSeq genes that were not mutated in at least three independent tumors were also eliminated.

RESULTS

1. GENERATION OF TRANSGENIC MICE CARRYING THE CONSTRUCTION K5-SB11

In order to obtain K5-SB11 transgenic mice, SB11 DNA was cloned into an expression vector that contained a 5.2 kbp fragment from the 5'-flaking region of the bovine keratin K5 gene, a β-globin intron and poly adenylation signals (**Figure 27**) (143). After that, the K5-SB11 DNA fragment of 8.4 kbp was microinjected in B6D2 F2 mice embryos. Several transgenic founder mice were generated in the Transgenic Mice Core Facility of CIEMAT, and a transgenic line was established by breeding with B6D2 F1 mice. These mice were viable and fertile, and did not show any overt phenotype.

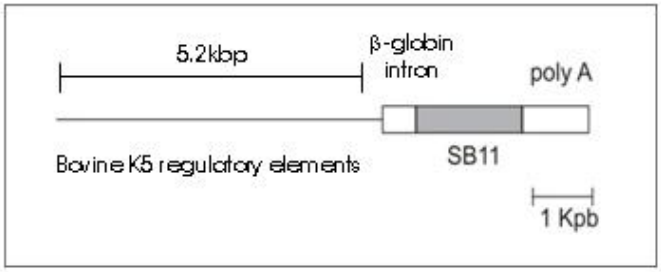

Figure 27. Representation of the K5-SB11transgene.

1.1. SB11 transposase expression follows the same pattern of keratin K5

In order to characterize the expression pattern of SB11 in K5-SB11 transgenic mice, we used Northern blot, Western blot and immunohistochemical techniques.

1.1.1. Analysis of transgene expression by Northern blot

The expression of SB11 in K5-SB11 mice was tested by Northern blot assays of RNA isolated from skin of transgenic and control mice (non-transgenic littermates). A band of the expected size, recognized by a SB transposase specific probe, was found only in transgenic mice samples **(Figure 28).** These results indicate that SB11 is expressed in the skin of K5-SB11 transgenic mice

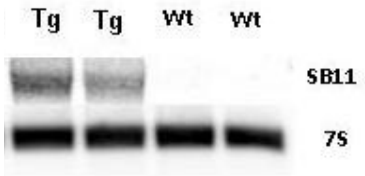

Figure 28. Northern blot analysis of skin mRNA isolated from K5-SB11 transgenic mice (**Tg**) and control mice (**Wt**)

1.1.2. Analysis of SB11 transposase protein by Western blot

With the aim of characterizing the expression pattern of the SB11 protein in K5-SB11 transgenic mice we performed Western blot analyses of protein extracts from control and transgenic mice. We analyzed several tissues and organs that contain either stratified epithelia or cells that express keratin K5.

Figure 29 shows high expression levels of SB11 protein in transgenic mice in tissues like foot sole, tail and back skin and considerable expression levels in palate, tongue, thymus and aglandular stomach (**Figure 29 a-c**), indicating that the expression pattern of SB11 in K5-SB11 mice roughly paralleled that of keratin K5.

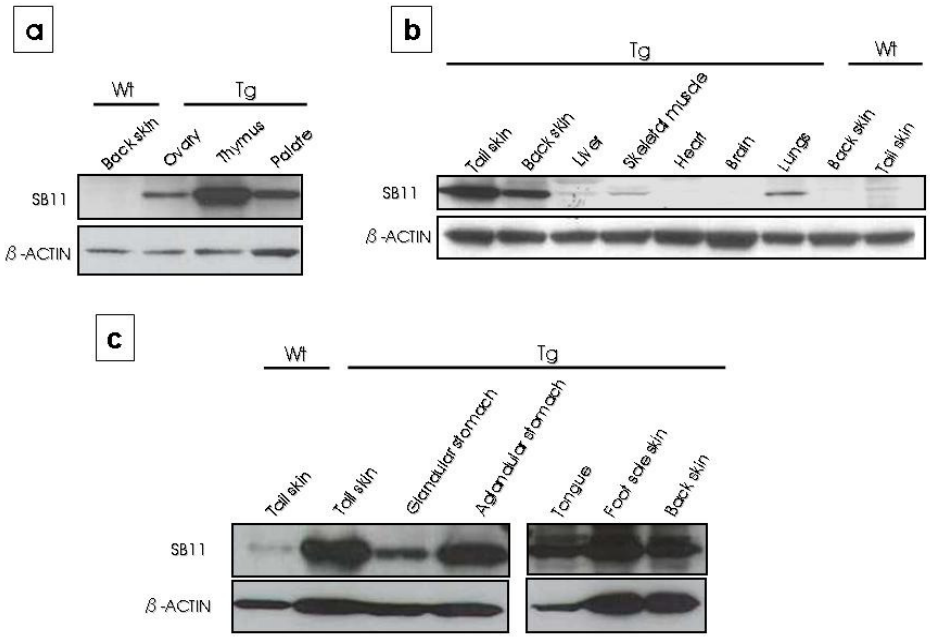

Figure 29. Western blot analysis of SB11 expression in several tissues from K5-SB11 transgenic mice (Tg) and control mice (Wt)

1.1.3. Immunohistochemical analysis of the SB11 transposase protein expression

In order to assess the expression pattern of the SB11 protein in K5-SB11 transgenic mice at cellular level, we performed an immunohistochemical assay using an antibody against SB transposase in organs and tissues where K5 is expressed (in particular, those with stratified epithelia, in myoepithelial cells of exocrine glands and in thymus stromal cells). The SB11 expression pattern was compared to the keratin K5 expression pattern by using an antibody specific for keratin K5. SB antibody staining was detected in most of the basal cells of stratified epithelia from back skin (**Figure 30b**), tail skin (**Figure 30e**), outer root sheath cells of hair follicles (**Figure 30h**),

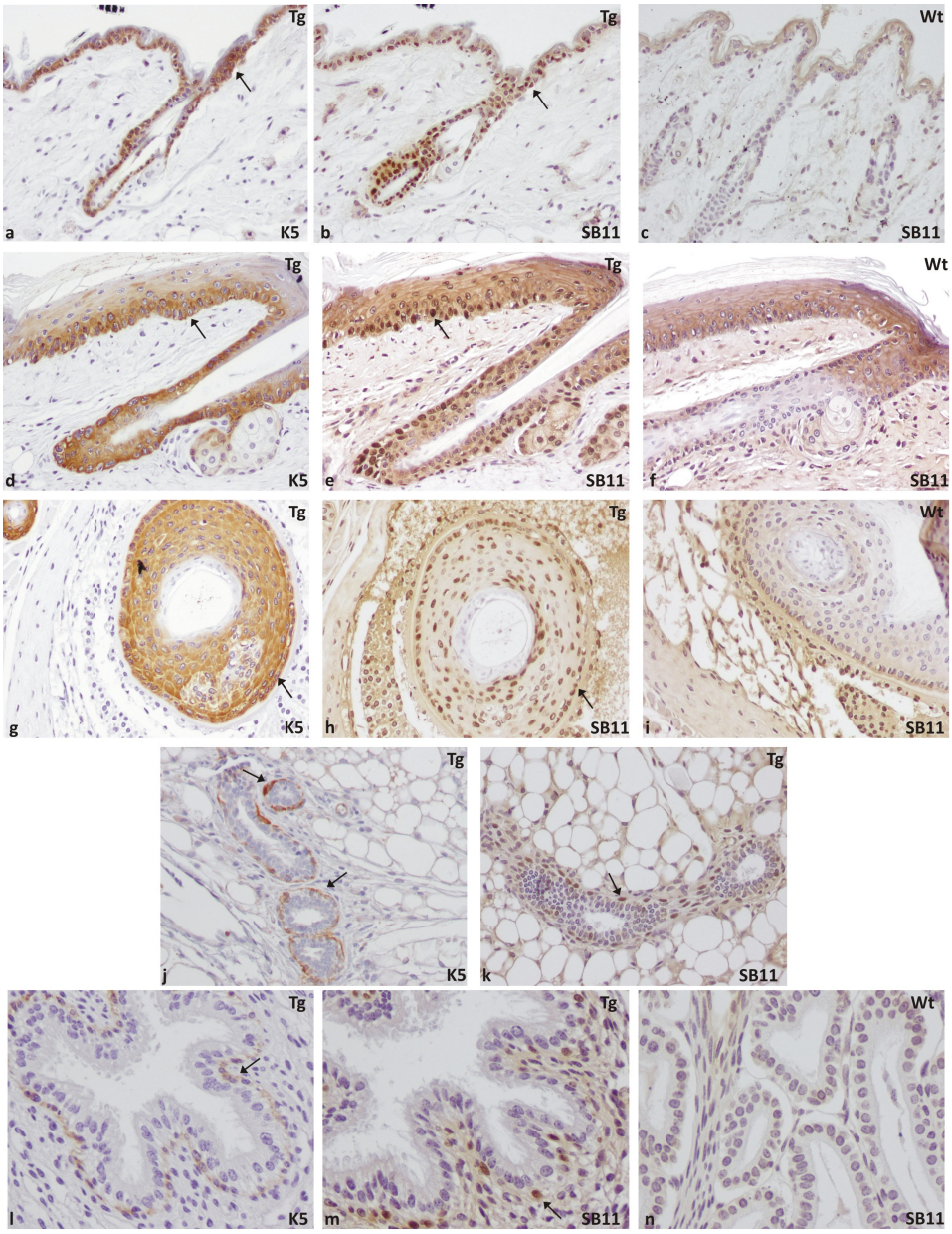

Figure 30. Immunohistochemical analysis of K5 and SB11 proteins. Histological sections from K5-SB11 transgenic (**Tg**) or control (**Wt**) mice were stained with K5 and SB antibodies, as indicated in back skin (**a-c**); tail skin (**d-f**); hair follicles (**g-i**); mammary gland (**j, k**) from a virgin female and prostate (**l-n**). The objectives used were: 20X (**a- k**) and 40X (**l-n**). Arrows indicate the K5 and SB11 expression in basal cells of the skin epithelia (**a**, **b**, **d** and **e**), outer root sheath of hair follicules (**g** and **h**), myoepthileial cells of mammary gland (**j**, **k**) and prostate (**l**, **m**)

oesophagus, oral epithelium, tongue, palate and foot sole (not shown), and in myoepithelial cells of exocrine glands from mammary gland (**Figure 30k**) and prostate (**Figure 30m**), salivary glands, pancreas and in thymic stromal cells (not shown). SB antibody gave no staining when assayed in tissues from non transgenic mice (**Figure, 30: c, f, i and n**), proving the specificity of the antibody used. SB11 expression pattern was similar to that of keratin K5 (**Figure 30:a, d,**

g, j and l) with the exception that K5 showed a cytoplasmic location, as expected, while SB11 was detected in the nucleus of the cells, due to the nuclear localization signal present in the transposases of the SB group.

1.1.4. Immunofluorescence and IHC analysis of the SB11 transposase protein expression in the hair follicles.

With the purpose of seeing if the SB11 was expressed in the epidermal stem cells, that are thought to be a cell population responsible of the origin of skin cancer among others, we decided to check if SB11 was expressed in the bulge region of the outer root sheath of the hair follicle, where these stem cells are located. Therefore, we used a CD34 antibody which is known to be a marker of the bulge stem cells (185) and a SB11 antibody.

Our immunoflourescence analysis showed that as expected, SB11 was co-expressed with CD34 in the immunofluorescence analysis in the bulge area of hair follicles (**Figure 31**). The same results were obtained by immunohistochemical analysis of the expression of CD34 and SB11 proteins.

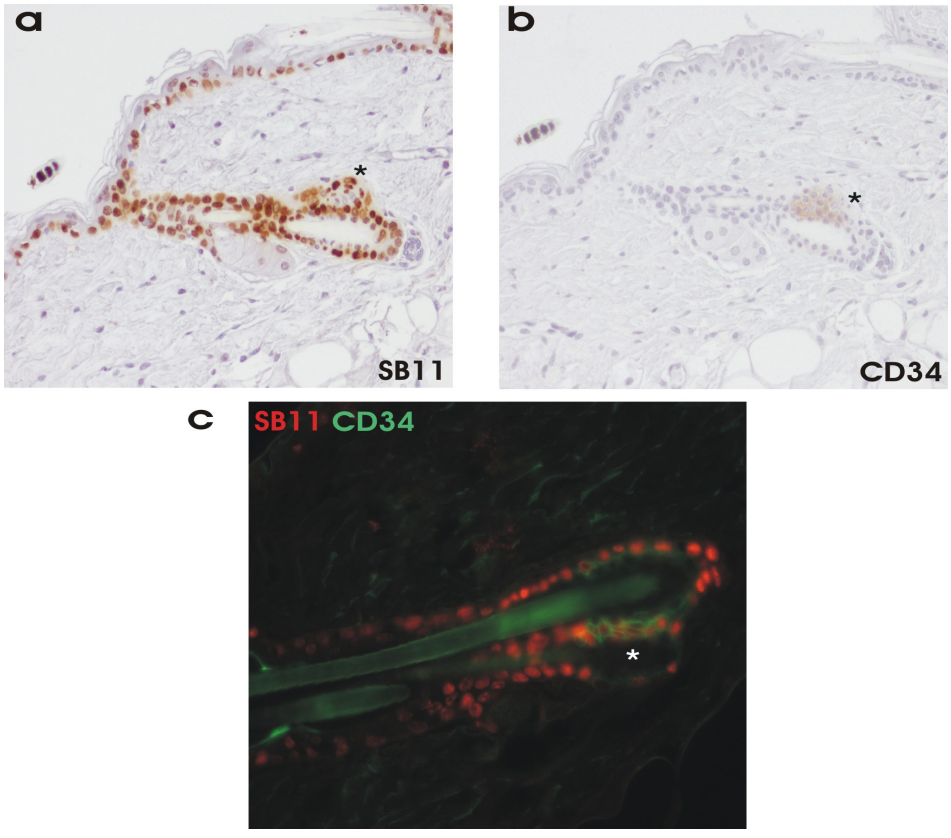

Figure 31. IHC and IF analysis of SB11 and CD34 proteins. Histological sections from K5-SB11 transgenic (Tg) mice were stained with SB11 and CD34 antibodies, as indicated in skin (**a**, **b**) and hair follicles (**c**). Asteriks indicate the SB11 and CD34 expression in cells from the bulge of the outer root sheat of hair follicules (**a**-**c**). The objectives used were: 20X (**a**-**c**) and 40X (**l**-**n**)

2. GENERATION OF DOUBLE TRANSGENIC MICE SB/T2 AND DETERMINATION OF TRANSPOSITION IN SKIN

2.1. Generation of the cohort of SB/T2 double transgenic mice

We wanted to obtain mice carrying transposons in their genome susceptible to be mobilized by the action of the SB11 transposase. With this aim, it was necessary to have the two transgenes (transposon and transposase) in the cells, so that transposons could be cut out from their original integration site, mobilized and pasted elsewhere in the genome of the cells expressing SB11 transposase. With this idea in mind, we decided to breed K5-SB11 mice (named as SB) with two different lines of transgenic mice carrying a concatemer of T2Onc2 transposons (named as T2)(50)**.** Both T2Onc2 transgenic lines differed in the original chromosomal localization of the transposon concatemer and in their copy number. It was estimated that line 6070 carries more than 200 copies located in chr4 and line 6113 carries around 350 copies in chr1(50). After breeding, we obtained a total number of 177 mice. As shown in **Figure 32**, all possible genotypes were obtained in these crosses. However, in line 6113, we found a decrease in the number of observed double transgenic mice in relation to the number expected ($p = 0.033$). This low proportion of SB/T2 double transgenic mice viable of the offspring was reported previously in similar experiments expressing SB11 ubiquitously(50), and it has been attributed to lethality induced by SB transposition and/or DNA damage not repaired after SB excision. Its frequency varies among the different T2Onc2 lines.

A

SB x T2(6070)	Observed	Expected
Non Tg	12	13.25
T2(6070)	8	13.25
SB	19	13.25
SBT2(6070)	14	13.25
Total	53	53

B

SBx T2(6113)	Observed	Expected
Non Tg	39	31
T2(6113)	37	31
SB	30	31
SBT2(6113)	18	31
Total	124	124

Figure 32. Mice genotype distribution. A) T2Onc2 6070 line. Genotype distribution presented a χ^2 value of 4.73 ($p = 0.192$). B) T2Onc2 6113 line. Genotype distribution presented a χ^2 value of 8.71 ($p = 0.033$) denoting that offspring distribution in T2Onc2 6113 line did not adjust to mendelian proportions.

2.2. Analysis of transposition events in SB/T2 double transgenic mice

To study if transposon mobilization was taking place in the double transgenic mice containing both the SB11 transposase and the T2Onc2 transgenes, we followed two different approaches.

2.2.1. PCR analysis of transposition

A PCR excision assay was done with the purpose of detecting transposition events within epidermal tissue of SB/T2 double transgenic mice. PCR analysis of the transposon array would result in a 2.2 kbp band in non-mobilized copies of the transposon and in a 225 bp band in case of transposition (**Figure 33a**, below)**.** The band of 225 bp, due to its smaller size, is preferably amplified when transposition is taking place in only some of the copies, and usually prevents the simultaneous presence of the 2.2 kbp band. So, the presence of the 225bp band indicates that at least some transposon copies have been excised. We found that the 225 bp band was amplified in all the double transgenic mice analyzed by PCR analysis of tail DNA (for an example, see (**Figure 33b**), proving that transposition was occurring in double transgenic mice. However, this type of analysis does not give a quantitative estimation of the transposition rate in double transgenic SB/T2 mice.

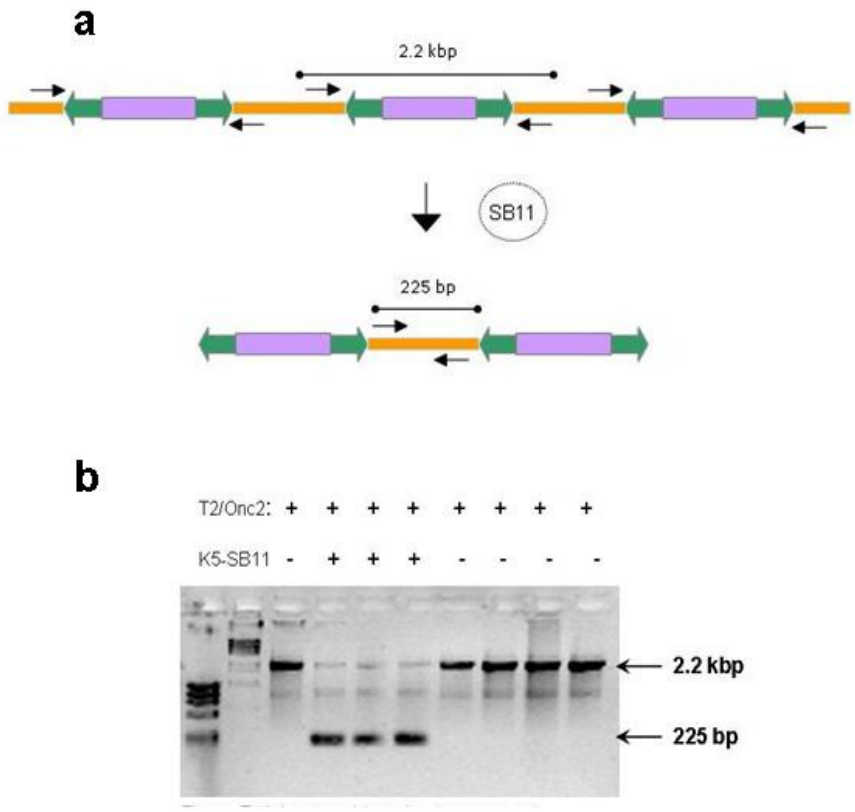

Figure 33. a) Primers amplify a 2.2 kbp product if a transposon has not been mobilized from the concatemer . If transposon excision occurs, a smaller band of 225bp is detected instead. **b**) Ethidium Bromide-stained agarose gel analysis of PCR products from tail tips DNAs of single and double Tg mice. The transgene(s) present in each mouse are indicated by **+** and **-** on the top. First and second lane corresponds to molecular weight markers.

2.2.2. Southern Blot analysis of transposition

The previous PCR excision approach indicated that transposition was taking place in the tails of SB/T2 mice, but did not allow a precise quantification of the transposition rate in each sample. Therefore, we decided to perform a Southern Blot analysis of purer cellular populations, using DNA isolated from the epidermis and dermis of double transgenic mice SB/T2.

Figure 34 shows how transposition can be detected by Southern Blot analysis. Bam HI sites are located within the plasmid sequences that flank each transposon in the concatemer and in the transposon itself. Digestion with Bam HI generates a 500 bp fragment that can be detected using a specific probe. When a copy of the transposon is excised from the concatemer and reintegrated into a new genomic location, Bam HI sites will not be flanking the transposon as in the concatemer. Consequently, it is expected that after transposition, BamHI fragments will have a different, variable and unknown size according to where the transposon copy integrates and where the nearest BamHI restriction site is placed. Therefore, we were able to recognize if excision of transposon copies from their original site has taken place because it will result in decreasing intensity (or even disappearance) of the 0.5 Kb band, and reinsertion into new sites in the genome may result in the detection of new very low intensity bands.

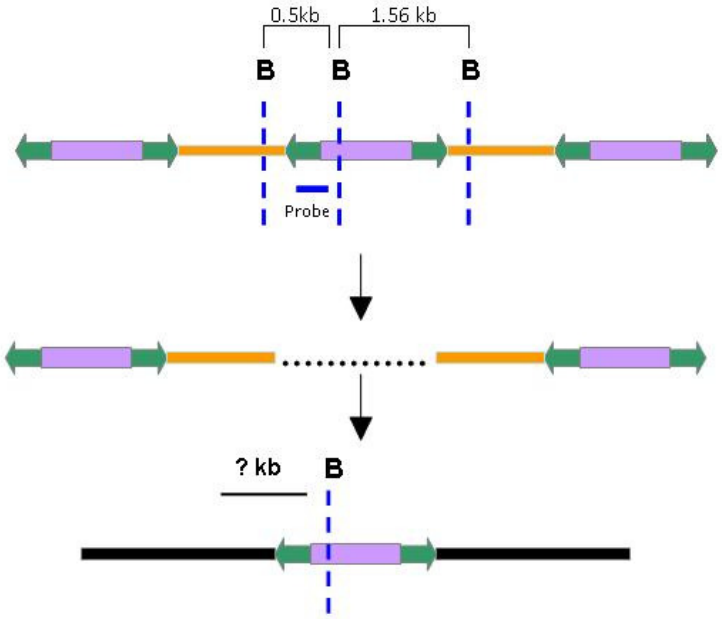

Figure 34. BamHI digestion generates a band of **500bp**, detected by a T2Onc2 probe when the transposons are located within the concatemer. Whenever the transposons are excised and reintegrated into a new position in the genome, the T2Onc2 probe detects BamH1 digestion fragments of unknown size. "B" stands for BamHI digestion sites.

DNA was isolated from samples of skin epidermis and dermis from double SB/T2 transgenic mice (which carried both transposons and the SB11 transgene), and from T2 transgenic mice

that did not carry the SB11 transgene, as non-transposed control samples. A decrease in the intensity of the 500bp band was detected in samples from SB/T2 mice, indicating that some copies of the transposon have been mobilized. This decrease is clearly observed in lanes 1, 4, 5, 7 and 10 in **Figure 35**, when compared to control samples (lanes 2, 3, 6, 8 and 9, with no transposition). Line 11 does not present the band of 500bp because it corresponds to a DNA sample from a non-transgenic mouse. This decrease of intensity in the 500bp band correlates, as expected, with the presence of a smear of DNA fragments of higher size that correspond to mobilized transposon copies when integrated into different parts of the genome (asterisks in **Figure 35**).

Although transposition events should not occur in the dermis, some transposon mobilization is observed in dermis samples from double transgenic mice (see lanes 1 and 4). This could be explained considering that hair follicles remain in the dermis after epidermal separation. As outer root sheath cells from hair follicle express transposase (**Figure 30b, 30e, 30h and 31c**), it is not surprising that transposition is also detected in these samples. **Figure 35** also suggests that the rate of transposition becomes higher with time, (from birth to adult mice), as there is a more pronounced decrease of intensity in the 500bp band in samples from older mice (compare lanes 7 and 10, that belong to the same animal at different ages). The quantification of this southern blot indicated that only around 10% of the transposon copies remained non-mobilized by the age of 8 weeks. So, we developed an animal model in which most of the hundreds of transposon copies were mobilized at the age of 8 weeks at least once. We next studied the mutagenic effects of this transposon mobilization in the development of skin cancer.

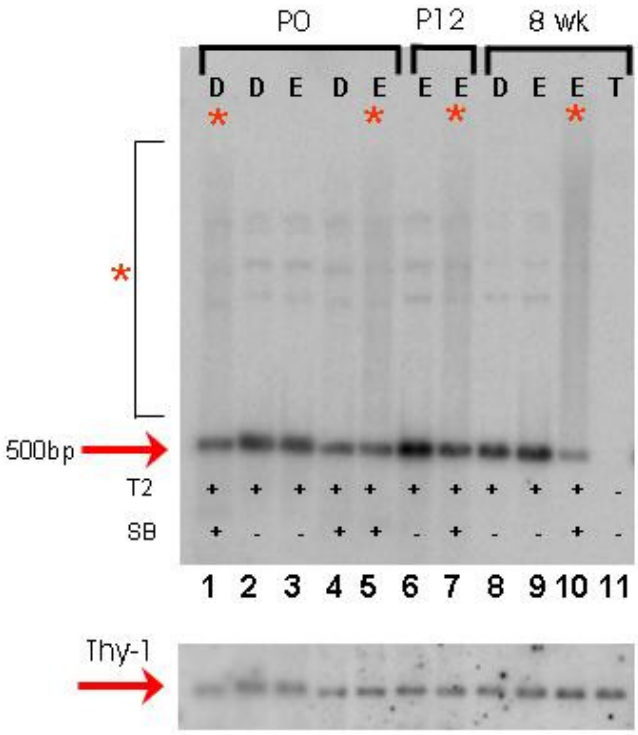

Figure 35. Southern Blot analysis of SB/T2 samples (+/+), T2 samples (+/-) and non-transgenic sample (-/-). *Thy-1* probe was used as a loading control. "**D**" and "**E**" corresponds to DNA isolated from dermis and epidermis samples respectively. Asterisks mark samples with a visible DNA smear. P0, P12 (postnatal) and 8 weeks indicate mice age.

3. EFFECT OF TRANSPOSON MOBILIZATION ON SKIN CANCER

3.1. SB/T2 mice do not show spontaneous skin tumours

These SB/T2 mice were viable, fertile and did not present phenotypical alterations, except for some alopecia and moderate lighter weight and smaller size.

In previous experiments done with transgenic mice carrying the transposon T2Onc2 and an ubiquitously expressed SB11 transposase, double transgenic mice developed tumors, mainly of hematological origin (33). However, in our experiment, double transgenic SB/T2 mice did not develop more tumors than control mice at the age of 6 months despite the mutagenic effect of transposition.

Therefore, we followed two different strategies with the purpose of minimizing the time of tumor development and maximizing the number of tumors obtained. The first approach consisted in the use of a well known two-stage chemical protocol for skin carcinogenesis, which is based on tumor initiation with the carcinogenic agent DMBA and promotion with TPA (44,

165, 215), administered as indicated in the materials and methods section. In addition, we performed a second approach based on the use of Tg.AC mice. These mice carry a transgene with an activated *v-Ha-Ras* oncogene, making them genetically predisposed to the formation of skin tumors. Therefore, these mice only need the application of a promoter agent, such as TPA, for skin tumor development.

3.2. Carcinogenesis assays in SB/T2 transgenic mice

Following the first strategy mentioned previously, DMBA and TPA were applied on nine mice with B6D2 hybrid background. Among them, five were SB/T2 and four of them lacked the SB11 transposase, thus acting as a control.

The second approach was carried out using seven SB/T2 mice bearing also the Tg.AC transgene (named as SB/T2/Tg.AC) and five control T2/TgAC mice lacking the K5-SB11 transgene. These animals were treated only with TPA due to their Tg.AC background. Both lines of T2Onc2 transgenic mice (6070 and 6113) were used in the experiment. Genotypes and mice treatments are detailed in the **Figure 36**.

Mice ID	Genotype	Treatment	T2 line
G191	SB/T2/TgAc	TPA	6070
G196	SB/T2/TgAc	TPA	6070
G201	SB/T2/TgAc	TPA	6070
G240	SB/T2/TgAc	TPA	6070
G391	SB/T2/TgAc	TPA	6113
G393	SB/T2/TgAc	TPA	6113
G494	SB/T2/TgAc	TPA	6113
G197	T2/TgAc	TPA	6070
G231	T2/TgAc	TPA	6070
G233	T2/TgAc	TPA	6070
G493	T2/TgAc	TPA	6113
G495	T2/TgAc	TPA	6113
G230	SB/T2	DMBA/TPA	6070
G447	SB/T2	DMBA/TPA	6070
G392	SB/T2	DMBA/TPA	6113
G420	SB/T2	DMBA/TPA	6113
G492	SB/T2	DMBA/TPA	6113
G194	T2	DMBA/TPA	6070
G221	T2	DMBA/TPA	6070
G441	T2	DMBA/TPA	6070
G390	T2	DMBA/TPA	6113

Figure 36. Genotype distribution of transgenic mice used in the skin carcinogenesis experiment

3.2.1. Tumor progression kinetics

Tumor progression was followed after treatment with DMBA/TPA or TPA, depending on the genetic background, up to 42 weeks. Appearance of skin tumors was monitored weekly. If the

animals became moribund, they were euthanized and tumor samples were collected for histopathological and genetic studies. Interestingly, tumor developed sooner and with higher multiplicity in SB/T2/Tg.AC mice when compared to SB/T2 mice. As we can see in **Figure 37B**, by 9 weeks, all SB/T2/Tg.AC mice had developed skin tumors and the number and size of the tumors increased with time and arised sooner than in Tg.AC mice without transposition **(Figure 38B)**

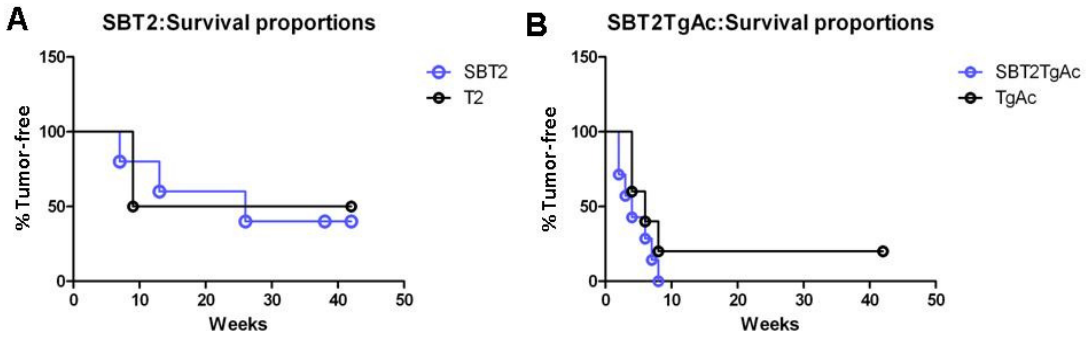

Figure 37. Tumor-free survival plots. **A**) Kaplan-Meier tumor-free survival analysis of SB/T2 vs. T2 control mice. **B**) Kaplan-Meier tumor-free survival analysis of SB/T2/Tg.AC vs. T2/Tg.AC control mice.

The average number of tumors per mice obtained was higher in mice with transposition (SB/T2/TgAC and SB/T2 mice) than in control groups that did not bear the transposase, (T2/Tg.Ac and T2, respectively) (**Figure 38**). Moreover, the latency period for tumor eruption was shorter compared to the control groups. In addition, papilloma regression phase occurred later in mice containing active transposons, starting between 22 and 24 weeks, than control mice, starting between 10 and 13 weeks (data shown in Supplementary Table 1).

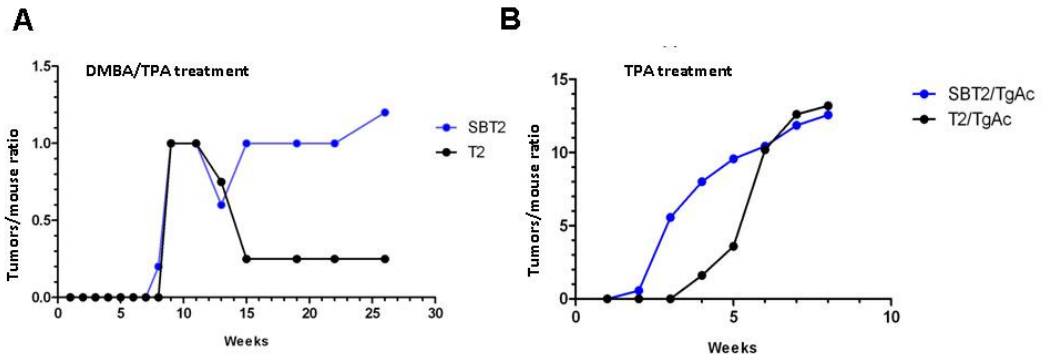

Figure 38. **Tumor-formation rate throughout carcinogenesis treatment. A**) SB/T2 and T2 mice. **B**) SB/T2/TgAc and T2/Tg.Ac mice. Data is included only up to 26 weeks in **A**) and up to 8 weeks in **B**) because at that point we started to sacrifice animals due to tumor number and size.

3.3. Histopathological characterization of the tumours

A number of 38 tumors corresponding to SB/T2 and SB/T2/TgAc backgrounds were analyzed in order to identify relevant histological features when compared to tumors isolated from control animals T2/Tg.Ac and T2 that did not carry SB11 transposase. In order to perform a more comprehensive histological analysis, we incorporated tumours isolated from SB/T2 and SB/T2/TgAc mice that were not involved in the carcinogenesis protocol. These tumors did not present histopathological features different from those generated in treated animals of the corresponding genotype and have been included within their same genotype groups for analysis. **(Figure 39)**

	Genetic background	Tumors
Transposition (38 tumors)	SB/T2/Tg.AC (32 tumors)	19 Papillomas (12 w/ basal proliferation) 7 SCC 3 BCCs 2 Kerathoacanthomas 1 Cervix Glandular Carcinoma
	SB/T2 (5 tumors)	1 Keratoacanthoma 1 Mammary Tubular Carcinoma 1 SCC 1 BCC + SCC 1 Mammary Adenocarcinoma
Control (10 tumors)	T2 (5 tumors)	1 Papilloma 2 Sebaceous Adenomas 2 Hyperproliferation
	T2/Tg.AC (5 tumors)	3 Papillomas (1 w/ basal proliferation) 1 Keratoacanthoma 1 Hiperkeratosis

Figure 39. Summary of tumor pathology analysis among the distinct genotype groups.

Most tumors corresponded to papillomas and were found mainly in back skin, but also in ear, eyelid and stomach. Despite the fact that we also found papillomas in control mice that did not bear the transposase SB11, tumors found in SB/T2 mice (both SB/T2/TgAC and SB/T2) presented a more malignant profile, characterized by basal cell proliferation and a tendency to invade the basal layers with infiltrating microcarcinoma foci. We also found several tumoral lesions that corresponded to SCCs (9) and BCCs (4), as well as 2 mammary gland carcinomas (**Figure 39**). In contrast, tumors obtained in control mice corresponded to benign papillomas and keratoacanthomas.

As mentioned above, papillomas found in SB/T2 mice presented several histopathological characteristics indicating a more malignant nature, such as basal cell hyperplasia and pronounced basal cell atypia (**Figure 40b, stars**) and microcarcinoma foci (**Figure 40d, asterisk and 40e-f, arrows**), accompanied with inflamation. Contrary to what happened in control animals (**Figure 40a),** SB/T2 mice also displayed BCCs with squamous differentiation **(Figure 40g, arrows)** and sebaceous differentiation foci represented by vacuolated clear cytoplasm of mature sebaceous cells (**Figure 40k-l, arrows**) and basal cell proliferation with atypia (**Figure 40h, stars**); and SCCs with infiltration of submucosa (**Figure 40i, star**) and accused atypia and anisokaryosis (**Figure 40j, stars**). All these alterations found in tumors in SB/T2 mice confirmed that they indeed exhibited a high tumorigenic potential that was not found in control mice tumors.

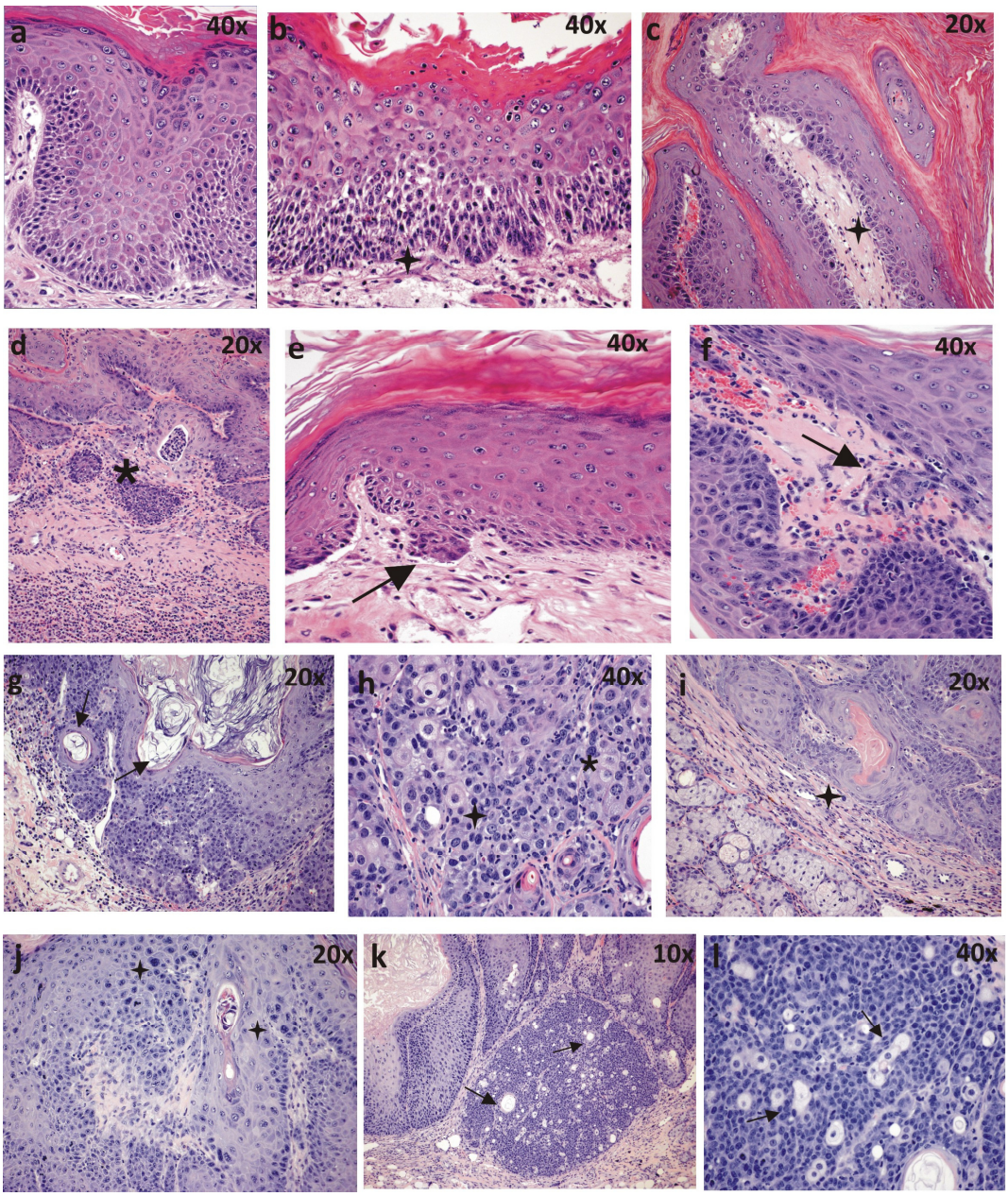

Figure 40. Hematoxylin and eosin stained histology sections from representative tumors in T2/Tg.AC (**a**) and SB/T2/Tg.AC mice (**b–i**) and SB/T2 mice (**j**-**l**). Squamous papilloma from control animal (**a**) and with basal proliferation (stars in **b**), basal cell atypia (stars in **c**) and microcarcinoma foci (arrows in **d**,-**f**). BCC samples presenting squamous differentiation (arrows in **g**) and sebaceous differentiation (arrows in **k** and **l**), anisokaryosis (asterisk in **h**), anisocytosis and atypia (star in **h**). SCC samples (**i**-**j**) with submucosa infiltration, anisokaryiosis (stars in **j**). **l** is a magnification of **k**.

4. ANALYSIS OF TRANSPOSON INSERTION SITES IN TUMORS GENERATED IN SB/T2 TRANSGENIC MICE

4.1. Some tumors have novel transposon insertions clonally amplified

A Southern Blot analysis of 20 tumour samples from 2 different mice (G196 and G391) and 2 DNA control samples isolated from tail skin was done with the aim of studying the mobilization of T2Onc2 transposon in skin tumors from SB/T2/Tg.AC transgenic mice. We used a specific probe that is able to recognize the T2Onc2 transposon transgene both when mobilized and non-mobilized from its original site in the concatemer (see **Figure 34**). Although some tumors present the same band pattern than non-tumoral DNA isolated from tail skin (**Figure 41,** lanes marked **C** and **C26**), in other tumor samples additional bands were present (for example, see red circles in **Figure 41**, lanes T12, T51 and T33,). These extra bands probably represent clonally amplified transposon insertions that are present in many of the tumoral cells, and thus could have been selected during tumor development. Theoretically, genes bearing these selected insertions are important in tumorigenesis.

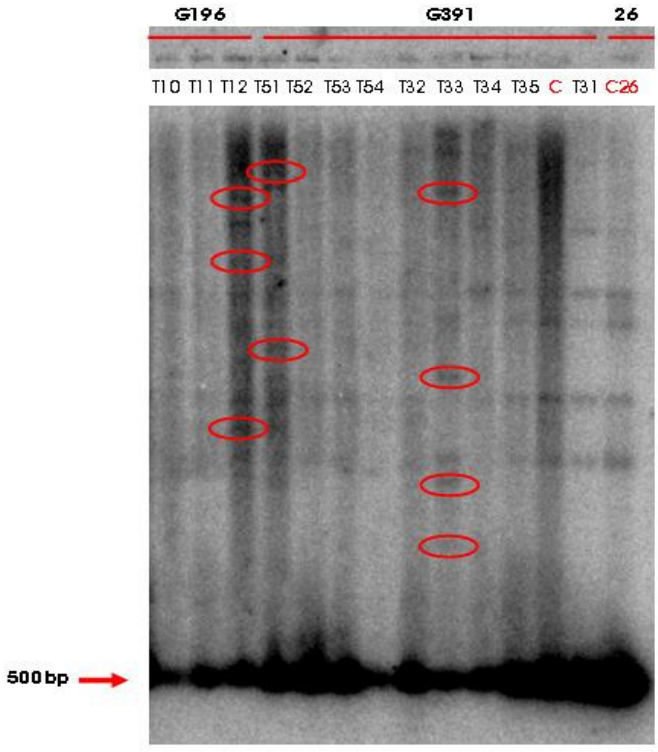

Figure 41. Southern Blot analysis of T2Onc2 integrations in tumor DNA isolated from "G196", "G391" and "26" SB/T2 mice. "**C**" and **C26** lanes correspond to isolated DNA from non tumoral tail skin samples.

4.2. Amplification of the transposon genomic DNA junctions

After having confirmed that transposition and clonal amplification of some integrations are taking place in tumors that developed in SB/T2 mice, we tried to identify the transposon insertion sites within tumor genomes, with the purpose of identifying putative genes affected by the integration of transposon copies. With this aim, DNA was isolated from tumors to sequence the junctions between T2Onc2 transposon and mouse genomic DNA, and therefore, to determine the transposon insertion sites.

Our initial approach consisted in PCR-amplification and shotgun cloning of the T2Onc2 insertion sites. Cloned sequences were subsequently sequenced and mapped to the mouse genome to localize the transposon-mouse genome junctions that allowed the determination of the genes affected by transposon insertion.

4.2.1. Shotgun cloning strategy

Genomic DNA was isolated from 6 skin papillomas corresponding to 3 SB/T2/Tg.AC mice, 1 SB/T2 mouse and from the skins of 2 control mice that did not carry active transposons. (**Figure 42**)

Mouse ID	Genotype	Treatment	Tumor ID	Sequenced insertions	Independent insertions	Hit on gene	No hit on gene	% Hit on gene	% No hit on gene
G196	SB/T2/Tg.AC	TPA	Tumor 1	4	4	3	1	75,0	25,0
G391	SB/T2/Tg.AC	TPA	Tumor 1	29	28	18	10	64,3	35,7
G391	SB/T2/Tg.AC	TPA	Tumor 2	16	7	6	1	85,7	14,3
G391	SB/T2/Tg.AC	TPA	Tumor 3	9	2	1	1	50,0	50,0
G213	SB/T2/Tg.AC	No treatment	Tumor 1	17	10	10	0	100,0	0,0
G392	SB/T2	DMBA/TPA	Tumor 1	6	3	2	1	66,7	33,3
		Total	**6**	**81**	**54**	**40**	**14**	**74,1**	**25,9**

Figure 42. Summary of sequenced transposon insertion sites found in tumors from SB/T2 mice

The DNA of each tumor was digested by restriction enzymes to split the left and right inverted repeats (IRL and IRR, respectively) of the transposon. These genomic fragments containing either the left or right part of the transposon were then amplified using a linker-mediated PCR modified protocol as described in Material and Methods. As a final result, we obtained a list of insertion sites from all tumours analyzed within the mouse genome. These results are summarized in **Figure 43**. In similar experiments performed in parallel using DNA of control mice without undergoing transposition, we did not obtain any insertion sites.

a

Mouse ID	Genotype	T2Onc2 line	Papillomas
G391	SB/T2/Tg.AC	6113 (chr 1)	T34, T33, T12
G196	SB/T2/Tg.AC	6070 (chr 4)	T1
G392	SB/T2	6113 (chr 1)	T1
G213	SB/T2/TgAC	6113 (chr 1)	T2

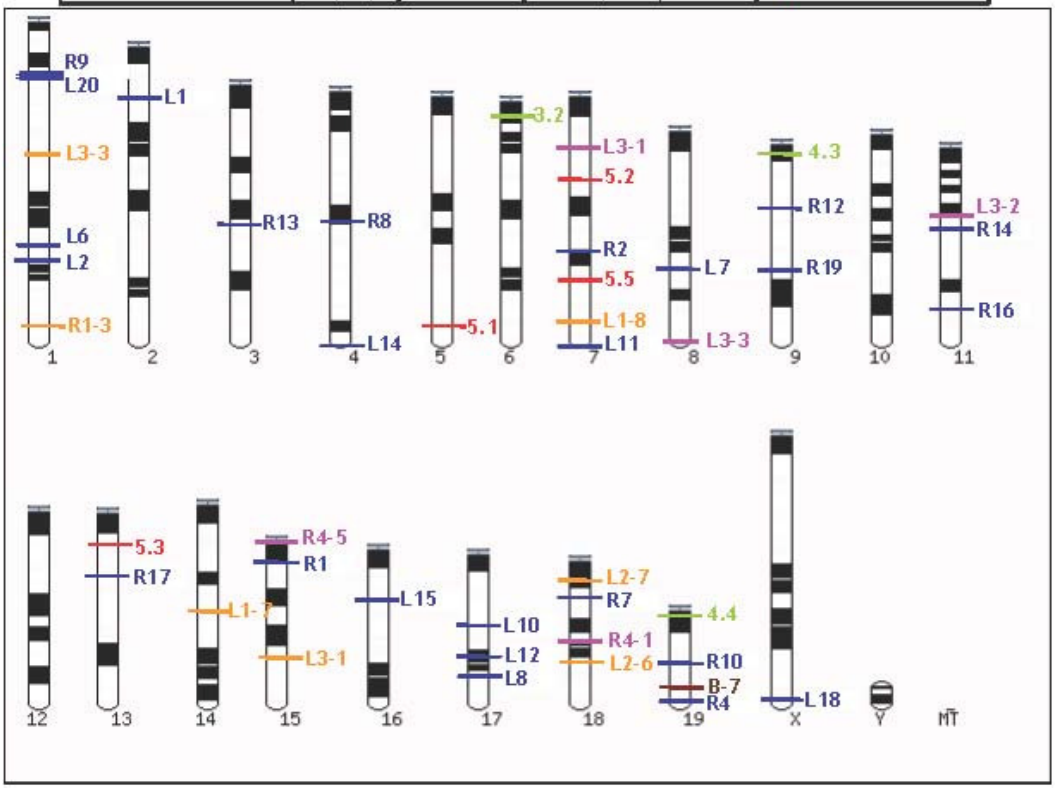

Figure 43. **a)** and **b)**. Chromosome map distribultion of transposon insertion sites. The position of the insertions of each tumor is marked by a colour line next to the insertion ID. Each color corresponds to the tumor ID color in **a)** table.

It is interesting to note that the rate of transposition near the transposon concatemer donor site (effect known as "local hopping") is low, as we do not detect much more integrations in Chr 1 or 4 than in the rest of the genome. This experiment confirmed that transposon mobilization was taking place within the tumors in animals containing transposon and transposase, supporting the results obtained in the Southern Blot analysis of the tumors shown in the previous paragraph. It also indicates the ability of the T2Onc2 transposon to move inside the genome, as transposon integrations were detected in all chromosomes (with the exception of the small Y chromosome). Another important data we found is that almost 75% of the sequenced insertions in the genome occurred in genes. This fact might indicate that these integrations in genes are being positively selected because they suppose a growth advantage to tumoral cells.

We did not find any CIS due to the limited number of sequences obtained using this approach, but this experiment confirmed that the transposon is able to integrate in any part of the genome and also that some insertions were clonally amplified .

However, this approach presented substantial limitations, such as the significant cost and labor intensiveness that were required to create a big shotgun library that have all possible transposon insertion sites cloned. Considering these limitations, we decided to use a pyrosequencing approach as an alternative. This technology eliminates the need to clone DNA fragments and the subsequent amplification and purification of templates before traditional capillary sequencing. It also allows sequencing the greatest possible number of transposon insertions within the genome, making possible the identification of enough transposon insertion sites to characterize CIS among the tumors.

4.2.2 Pyrosequencing

Genomic DNA was isolated from 76 independent tumors from 20 double transgenic mice containing transposon and transposase (**Figure 44,** and detailed list in supplementary **Table 2**)

Genotype and treatment	Number of mice	Number of tumors
SB/T2 (DMBA/TPA)	3	5
SB/T2/TgAC (TPA)	10	30
SB/T2/TgAC (none)	5	41

Figure 44. Summary of tumors used for DNA sequencing. Mice genotype and carcinogenic treatment are indicated

Samples were prepared for pyrosequencing as described in Materials and Methods section. An aliquot of every secondary PCR product was analyzed by agarose gel electrophoresis to verify the quality of the sample before performing the pyrosequencing run. (**Figure 45**)

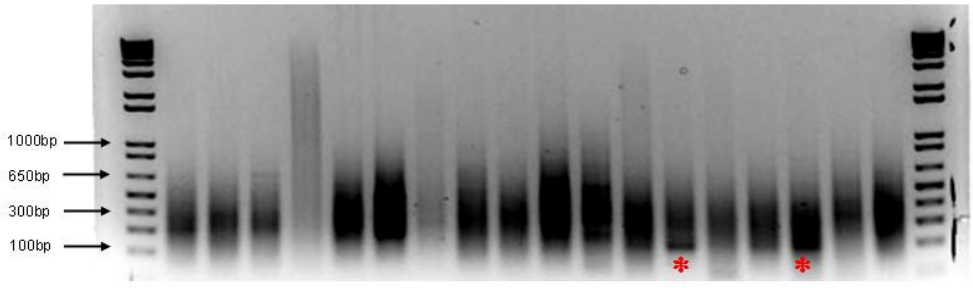

Figure 45. Example of a set of tumour simples analyzed on gel electroforesis after BfaI digestion and subsequent LM-PCR. Products appear as a low molecular weight smear. Some samples have more abundant junction products of a specific size that appear as clear bands (see **asterisks**)

A pool of all the resulting mixed barcoded PCR products was then sequenced using a 454 GS FLX platform

4.3. Common Insertion Sites (CIS) found within the genome allowed identification of skin cancer-related genes

GS FLX sequencing data output consisted of two different files containing all the generated sequence reads from the 79 analyzed tumours. Each file corresponds either to the left or right inverted repeats reads (IRL and IRR) and some genomic DNA flanking the transposon insertion sites. We obtained a total number of 550589 reads from IRL side and 563865 reads from IRR side. The mean size of the reads was 100bp.

Using the IAS (integration analysis system) software from the University of Iowa (http://ias.eng.uiowa.edu/IAS/), these sequence reads were analyzed to identify the transposon/linker sequences within the reads using BLAST and the flanking genomic insert was extracted into a separate file. After that, sequences were sorted by barcode identifier and the genomic junctions were mapped against the mouse genome (version mm9 of the database of the Genome Bioinformatics Group of the University of California Santa Cruz) using BLAT. Transposon integrations were aggregated for each tumor identifier (ID) and annotated based on the chromosome, base pair position and the number of sequence reads from the integration site. The UCSC genome annotation database was used to determine if integrations occurred within, or nearby a gene. If so, gene symbol, relative location within the gene, expected effect on the expression and orientation with respect to the gene were annotated. The orientation of the transposon in the insertion site reveals the presumed effect of the mutation over the expression of the gene (disrupted of promoted).

After finishing this process, two annotation files were generated: the IRR file and the IRL file, containing 44120 and 67952 mapped sequences, respectively. After that, both files were merged on one worksheet, so that the independent reads generated from transposon IRL and IRR sides that matched the same transposon insertion site could be combined. After condensation, a set of 108466 non redundant transposon integration sites from all tumors was generated, with tumors containing an average number of 1506 unique integration sites per tumor. Comparing the number of sequence reads and the number of mapped integration sites per tumor, we saw that in most tumors a large number of sequence reads resulted in a small number of integration sites (**Figure 46**). This can be explained by the fact that frequently sequenced sites represent clonally expanded transposon integration events in tumor cells that correspond to a small number of integration sites.

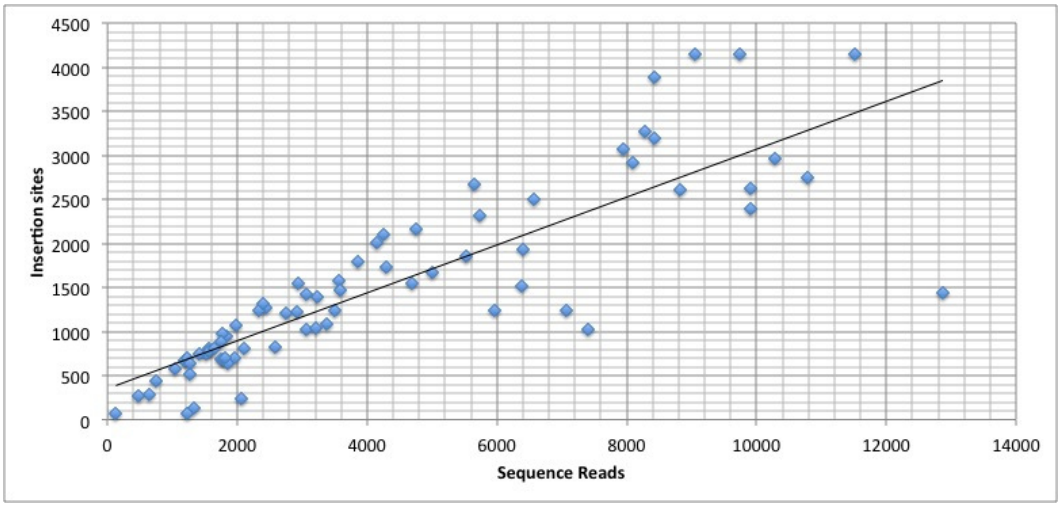

Figure 46. Comparison of the number of mapped sequence reads per tumor with the number of unique transposon insertion sites per tumor

The integration sites in the donor chromosomes, chr1 in T2Onc2 line 6113 and chr4 in T2Onc2 line 6070 were then removed to avoid false positive insertion events due to local hopping. Insertions corresponding to *En2* and *Foxf2* splice acceptors from the transposon were also removed, leaving a final number of 96768 non redundant insertion sites for subsequent clonal insertion sites analysis. A clonality score parameter for each integration site in each tumour was calculated considering the number of reads per insertion site and the number of reads at the top hit and a threshold value of 0.1 was applied to select the most significant clonal insertion events for subsequent CIS analysis.

A modified Monte Carlo statistic tool and a gene-centric method (gCIS) were used to identify the CIS regions in the list of tumor insertion sites. Both methods assume that background mutations will happen at random throughout the genome and that causative mutations will cluster in regions within or near cancer genes. These regions (CISs) would have a higher mutation rate than expected by chance. The Monte Carlo method considers that transposons will be randomly distributed at TA dinucleotide sites throughout the genome and creates a simulated integration pattern that can be used to calculate the significance of any candidate CIS genomic region by determining the probability that transposon could integrate by chance within a DNA fragment of a given size. Parameters such as the number of TA sites in the genome, the number of insertion events per tumour and the number of tumours are used in this simulation to identify those loci with more integrations than expected by chance. The gCIS method specifically examined transcribed regions of the genome, avoiding the identification of CISs that were too small or too large to be biologically meaningful.

Following this methodology, we finally obtained a list of 126 CIS **(Figure 47).**

Gene	# Tumors	p-value	Predicted Mutagenic effect
Notch1	32	0	Loss of function
Pard3	8	7,53E-15	Loss of function
Nsd1	7	1,79E-53	Loss of function
Cd44	6	9,05E-51	Gain of function
Narg1	5	2,24E-42	Gain of function
Wac	5	1,50E-38	Loss of function
Larp5	5	1,73E-36	Gain of function
Trp63	5	1,70E-32	Loss of function
Tnrc6b	5	5,15E-14	Loss of function
Trps1	5	3,86E-11	Loss of function
Syncrip	4	1,82E-43	Loss of function
Atp2a2	4	6,71E-40	Loss of function
Fat1	4	7,94E-22	Loss of function
Adam10	4	6,84E-17	Loss of function
1700081L11Rik	4	3,96E-16	Loss of function
Ankrd11	4	4,80E-15	Gain of function
Cdkn2a	3	4,91E-44	Loss of function
Tle3	3	2,59E-35	Loss of function
Maea	3	1,18E-30	Gain of function
Hnrnpm	3	2,88E-24	Loss of function
Fat2	3	2,59E-23	Loss of function
Ctnnd1	3	5,13E-21	Gain of function
Dhx15	3	8,75E-21	Gain of function
Ppp2r2a	3	4,05E-18	Loss of function
Pparg	3	1,36E-14	Gain of function
Ncor2	3	4,36E-14	Gain of function
Adra1a	3	1,41E-13	Gain of function
Bptf	3	1,73E-13	Loss of function
Rnf111	3	3,40E-13	Loss of function
Senp6	3	1,24E-12	Loss of function
Atf7	3	2,85E-12	Loss of function
Notch2	3	3,37E-11	Loss of function
Arih1	3	4,95E-10	Loss of function
Zmiz1	3	5,35E-10	Gain of function
Ehmt1	3	8,62E-10	Loss of function
Crebbp	3	4,18E-09	Gain of function
Atrx	3	8,34E-08	Loss of function
Prkch	3	9,17E-08	Loss of function
Hipk2	3	2,05E-07	Loss of function
Kdm6a	3	2,12E-07	Loss of function
Mef2a	3	2,71E-07	Loss of function
Gsk3b	3	4,97E-07	Loss of function
Cblb	3	1,68E-06	Loss of function
Eras	2	3,10E-53	Gain of function
Hspa5	2	4,49E-46	Loss of function
Trp53	2	1,30E-35	Loss of function
Oaf	2	1,33E-30	Loss of function
Sf1	2	1,33E-28	Gain of function
Cyp4a14	2	1,63E-28	Gain of function
Rab7	2	1,63E-28	Gain of function
Wdr79	2	1,81E-28	Gain of function
Erbb2	2	4,99E-27	Loss of function
Lsp1	2	4,49E-26	Gain of function
Ddx3x	2	4,17E-24	Gain of function
Snrp70	2	6,88E-24	Loss of function
Ripk4	2	3,02E-23	Loss of function
Rnf4	2	2,51E-22	Loss of function
Ddx17	2	2,92E-22	Gain of function
0610007L01Rik	2	3,86E-21	Gain of function
Atp1a1	2	2,39E-20	Loss of function
Clk4	2	7,66E-20	Loss of function
Sbno2	2	9,29E-20	Gain of function
Cd9	2	1,93E-19	Loss of function

Nfs1	2	2,02E-19	Loss of function
4931406C07Rik	2	5,94E-19	Loss of function
Nisch	2	7,39E-19	Gain of function
Ywhaz	2	7,55E-19	Gain of function
Cpn1	2	1,11E-18	Gain of function
Sh3bp2	2	2,27E-18	Gain of function
Ociad1	2	1,38E-16	Gain of function
Stat5b	2	9,01E-15	Loss of function
Rhoa	2	2,34E-14	Loss of function
B4galt5	2	4,65E-14	Loss of function
Pcf11	2	5,88E-14	Loss of function
Glt8d3	2	8,89E-14	Loss of function
Hnrnpc	2	1,04E-13	Loss of function
Atp1b3	2	1,05E-13	Loss of function
Ehf	2	1,26E-13	Loss of function
BC037393	2	2,46E-13	Loss of function
Zfp143	2	3,81E-13	Loss of function
Tnrc18	2	4,49E-13	Loss of function
Eif4e	2	4,62E-13	Loss of function
Matr3	2	1,94E-12	Loss of function
Csnk2a2	2	2,57E-12	Loss of function
Csnk1a1	2	5,00E-12	Loss of function
Rbm39	2	1,05E-11	Loss of function
Eif4enif1	2	1,55E-11	Loss of function
Paxip1	2	2,86E-11	Loss of function
Csde1	2	6,59E-11	Gain of function
Rxra	2	6,82E-11	Loss of function
Mar7	2	2,39E-10	Loss of function
Sfi1	2	4,33E-10	Loss of function
Dag1	2	4,49E-10	Gain of function
Actr3	2	5,16E-10	Loss of function
Hif1a	2	6,54E-10	Gain of function
Amotl1	2	1,67E-09	Loss of function
Herpud2	2	1,80E-09	Gain of function
Osmr	2	3,71E-09	Loss of function
Lin54	2	6,90E-09	Gain of function
Prkci	2	8,26E-09	Loss of function
Rtn4	2	1,17E-08	Loss of function
Tbca	2	1,44E-08	Loss of function
Ahnak	2	1,51E-08	Loss of function
Ubap2l	2	1,53E-08	Loss of function
Pafah1b1	2	1,72E-08	Loss of function
Tmod3	2	2,66E-08	Gain of function
Frmd6	2	3,93E-08	Loss of function
Ryk	2	4,48E-08	Gain of function
Stk24	2	5,34E-08	Gain of function
E2f3	2	5,70E-08	Gain of function
Polk	2	1,21E-07	Gain of function
Iqgap1	2	3,12E-07	Gain of function
Hc	2	3,27E-07	Loss of function
Uxs1	2	3,28E-07	Loss of function
Bcl11a	2	3,50E-07	Loss of function
Chd1	2	3,81E-07	Gain of function
Nfyc	2	4,04E-07	Gain of function
Tcerg1	2	5,02E-07	Gain of function
Cables1	2	6,63E-07	Loss of function
Ctdspl	2	7,68E-07	Loss of function
Mark3	2	8,82E-07	Loss of function
Pten	2	9,05E-07	Loss of function
Cul1	2	1,15E-06	Loss of function
Epc1	2	1,22E-06	Loss of function
Tmem45a	2	1,50E-06	Gain of function
Gfod1	2	1,62E-06	Gain of function

Figure 47. List of the 126 CIS genes ordered first by the number of tumors that share a same CIS and then by its statistical relevance

5. FUNCTIONAL ANALYSIS

Functional analysis of the 126 CIS list was done using the DAVID functional annotation tool (http://david.abcc.ncifcrf.gov/). The enriched Gene Ontology (GO) biological process terms obtained were related to transcriptional activity, cell proliferation, regulation of programmed cell death and epithelium development. Among GO molecular function terms, chromatin binding, transcription regulator activity and nucleotide binding were enriched.

These results suggest that there is a positive selection in tumor cells for transposon integrations in genes that have a role in transcription and in gene proliferation, processes that are known to be deregulated in tumorigenesis. Moreover, GO Pathway analysis showed that several transposon-mutated genes are involved in well-known cancer pathways, like *Wnt*, *p53* and *Notch* signalling pathways. In addition, a study of the protein domains present in the transposon-mutated genes showed enrichment in domains related to functions like protein kinase and serine/threonine protein kinases, also known to be altered in tumorigenesis (**Figure 48**).

The 126 CIS gene list obtained from mouse tumors was also confronted to the human genes listed in the human Cancer Gene Census (v15/03/2012). We found that 13 of the 126 genes (10%) had a homologue human gene (*ATRX, BCL11A, CBLB, CDKN2A, CREBBP, ERBB2, KDM6A, PPARG, NOTCH1, NOTCH2, NSD1, PTEN* and *TP53*), whose mutations have been catalogued to be causally implicated in cancer, so these data supports the idea that genes in the 126 CIS list may have a role in human carcinogenesis.

We also compared the 126 CIS to the human Catalogue Of Somatic Mutations in Cancer (COSMIC v59), that contains curated data from known cancer genes literature and systematic screens. In this analysis, we found that *NOTCH1, CDKN2A*, and *PTEN* have been identified as somatically mutated in 48 %, 13.2 % and 5.1 %, respectively, of human NMSC tumor samples, supporting the idea that genes in the CIS list might be important in human skin tumorigenesis.

TERM	BIOLOGICAL PROCESS	#GENES	*P*-VALUE
GO:0006350	Regulation of transcription	38	3.33E-06
GO:0008284	Regulation of cell proliferation	15	5.28E-05
GO:0032989	Cellular component morphogenesis	11	3.26E-04
GO:0060429	Epithelium development	9	0.001
GO:0012502	Regulation of programmed cell death	11	0.009
TERM	MOLECULAR FUNCTION	#GENES	*P*-VALUE
GO:000368	Chromatin binding	9	2.86E-05
GO:0030528	Transcription regulator activity	24	4.56E-05
GO:0000166	Nucleotide binding	32	4.63E-04
GO:0008270	Zinc ion binding	28	0.005
TERM	CELLULAR COMPONENT	#GENES	*P*-VALUE
GO:0031981	Nuclear lumen	21	1.08E-06
GO:0005681	Spliceosome	7	1.82E-04
GO:0031252	Cell leading edge	6	9.27E-04
GO:0030027	Lamellipodium	4	0.009
TERM	KEGG PATHWAYS	#GENES	*P*-VALUE
mmu05200	Pathways in cancer	13	0.0000272
mmu04520	Adherens junction	7	0.000062
mmu04530	Tight junction	8	0.000213
mmu04110	Cell cycle	7	0.001
mmu05215	Prostate cancer	6	0.001
mmu05223	Non-small lung cancer	5	0.001
mmu04310	Wnt signalling pathway	7	0.002
mmu05220	Chronic myeloid leukemia	5	0.005
mmu05219	Bladder cancer	4	0.007
TERM	PANTHER PATHWAYS	#GENES	*P*-VALUE
P00059	p53 pathway	5	0.004
P0045	Notch signaling pathway	7	0.005
P04398	p53 feedback loops2	5	0.008

Figure 48. GO annotation terms of the 126 CIS list. % refers to the percentage of the genes from the 126 CIS gene list that were significantly enriched in a specific GO category

5.1. Selection of 16 genes from the 126 CIS list for further analysis

We selected for further study a list of 16 genes that were mutated in at least in 4 different mice tumors from more than one mouse (**Figure 49**).

Gene symbol	Gene Name	p-value	# of tumors	# mice
Notch1	Notch homolog 1, translocation-associated (Drosophila)	0	32	13
Pard3	Par-3 partitioning defective 3 homolog (C. elegans)	7,53E-15	8	6
Nsd1	Nuclear receptor binding SET domain protein 1	1,79E-53	7	4
Cd44	Cd44 antigen	9,05E-51	6	5
Naa15	N(alpha)-acetyltransferase 15, NatA auxiliary subunit	2,24E-42	5	5
Wac	WW domain containing adaptor with coiled-coil	1,50E-38	5	5
Larp5	La ribonucleoprotein domain family, member 4B	1,73E-36	5	2
Trp63	Tumor protein 63	1,70E-32	5	3
Tnrc6b	Trinucleotide repeat containing 6B	5,15E-14	5	4
Trps1	Trichorhinophalangeal syndrome 1	3,86E-11	5	5
Syncrip	Synaptotagmin binding, cytoplasmic RNA interacting protein	1,82E-43	4	4
Atp2a2	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	6,71E-40	4	3
Fat1	FAT tumor suppressor homolog 1 (Drosophila)	7,94E-22	4	4
Adam10	ADAM metallopeptidase domain 10	6,84E-17	4	4
1700081L11Rik	RIKEN cDNA 1700081L11 gene	3,96E-16	4	5
Ankrd11	ankyrin repeat domain 11; hypothetical protein LOC100128265	4,80E-15	4	4

Figure 49. **Genes from the 126 CIS list mutated in at least 4 different tumors**. *P-value* refers to the significance of each gene contained in the original 126 CIS list.

Despite the fact that we only found a few genes from the 126 CIS list that were somatically mutated in human skin cancer samples present in the COSMIC catalogue, we observed that somatic mutations in 15 out of this 16 gene list (**Figure 50**) have been detected in tumors originated in tissues other than skin.(**Figure 50**). Moreover, somatic mutations were found in 7 genes out of the 16 gene list in human SCC samples: *NOTCH1, NSD1, SYNCRYP, TP63, KIAA1267, PARD3* and *TRPS1*, strengthening the idea that these genes could be relevant in human skin carcinogenesis.

Gene Symbol	Total Samples	Mutations found	Skin samples	Mutations found in skin*	Tissue
Notch1	6044	760	117	15	663 hematopoietic/lymphoid, 32 SCC upper aerodigestive tract, 16 lung carcinoma (several types), 12 glioma, 14 breast carcinoma, 3 serous carcinoma ovary, 1 intestine adenocarcinoma, 2 SCC oesophagus, 1 ductal carcinoma pancreas,15 SCC non-melanoma skin
Nsd1	201	16	2	2	7 SCC upper aerodigestive tract, 2 serous carcinoma ovary,2 hematopoietic/lymphoid, 1 lobular breast carcinoma, 1 kidney carcinoma, 2 serous carcinoma ovary, 2 SCC non-melanoma skin,
Cd44	219	1	6	0	1 pancreas
Syncrip	97	7	2	2	2 breast carcinoma, 2 serous ovary carcinoma, 2 skin non- melanoma SCC, 1 serous carcinoma ovary, 1 SCC upper aerodigestive tract
NAA15	95	4	1	0	1 SCC upper aerodigestive tract, 1 serous carcinoma ovary, 1 malignant melanoma
Atp2a2	95	5	1	0	1 hematopoietic/lymphoma, 3 serous carcinoma ovary, 1 malignant melanoma
Wac	157	4	0	0	1 hematopoietic/lymphoma, 1 serous carcinoma ovary, 1 adenocarcinoma colon
Larp4b	93	4	0	0	1 breast carcinoma, 2 serous carcinoma ovary, 1 astrocitoma
Tp63	417	16	8	1	1 breast carcinoma,6 SCC upper aerodigestive tract, 2 serous carcinoma ovary, 2 malignant melanoma skin, 4 adenocarcinoma lung, 1 SCC non-melanoma
Fat1	47	25	0	0	9 SCC upper aerodigestive tract, 13 serous carcinoma ovary, 1 glioma, 1 pancreas, 1 hematopoietic/lymphoma
Adam10	247	4	34	0	1 serous carcinoma ovary, 3 malignant melanoma
KIAA1267	76	7	3	1	1 ductal carcinoma, 1 hematopoietic/lymphatic, 1 SCC upper aerodigestive tract, 1 oligodendroglioma, 2 malignant melanoma, 1 skin SCC non-melanoma
Ankrd11	142	9	0	0	6 serous carcinoma ovary, 1 astrocytoma, 1 adenocarcinoma colon, 1 adenocarcinoma stomach, 2 SCC upper aerodigestive tract
Pard3	99	6	1	1	2 breast carcinoma, 2 SCC upper aerodigestive tract, 1 serous carcinoma ovary, 1 SCC non-melanoma skin
Tnrc6b	22	0	0	0	
Trps1	224	11	9	3	1 breast ductal carcinoma, 3 serous carcinoma ovary, 1 ductal carcinoma pancreas, 1 nonsmall cell carcinoma lung, 1 clear cell carcinoma kidney, 2 malignant melanoma, 3 SCC non-melanoma skin

Figure 50. Summary of somatic mutations found in the list of 16 genes selected from the 126 CIS list using the COSMIC database for analysis (COSMIC v59). *Mutations in skin excluding melanomas

6. STUDY OF THE EXPRESSION OF SELECTED GENES IN A SET OF HUMAN SKIN TUMORS SUGGEST THAT THEY CAN HAVE A ROLE IN HUMAN SKIN TUMORIGENESIS.

We then studied if mutations in any of these 16 genes are also important in the origin and development of human skin cancer. Our first approach consisted in the study of the expression levels of these genes in human skin cancer samples. We selected a reduced set of 11 genes for expression analysis because of its biological relevance in human skin carcinogenesis based on previous studies and also because the rest of the genes that were not selected for the study presented several technical drawbacks for gene and protein expression studies due to the large number of splicing forms or lack of available antibodies. Selected genes were *NOTCH, PARD3, NSD1, WAC, TNRC6B, TRPS1, ATP2A2, FAT, ADAM10, SYNCRYP* and *TP63*. A summary of the relevant biological features that link these genes to human skin cancer is presented below:

-*NOTCH1* encodes a transmembrane protein characterized for its role in development. Members of the Notch family belong to an evolutionary conserved signalling pathway implicated in different functions during embryogenesis and self-renewing of tissues. These functions include the maintenance of stem cells, cell fate specification, proliferation and apoptosis. Loss of Notch1 function has been correlated to development of BCCs and SCCs in mouse models of epidermal carcinogenesis (129), (139), and recently, mutations that disrupt this gene have been found in several forms of human SCCs(200), (1).

-*PARD3* encodes a protein that is involved in cell polarity and has been shown to be downregulated in colorectal tumour samples (116), in esophageal SCCs cell lines (216), and in ductal breast cancer samples (184). It is also somatically mutated in cutaneous SCCs (54). A non -functional PARD3 protein is related to a loss of polarized phenotype, epithelial cell hyperproliferation and acquisition of invasive potential (152).

-*NSD1* encodes a transcriptional regulator protein with H3K36 methyltransferase activity, being able to modify the activity of target genes. Nsd1 has been found to be altered in multiple myelomas, lung cancer, neuroblastomas and glioblastomas. Recently, somatic mutations have been found in some cases of skin SCCs (176). In addition, mutations in the *NSD1* gene are associated with Sotos overgrowth syndrome, a dominant autosomal condition caused by loss-of-function mutations or deletions and associated haploinsufficiency of the *NSD1* gene. This disease is characterized by overgrowth and macrocephaly, together with an increased risk of tumorigenesis in several locations, including skin (59).

-*TRPS1* encodes a GATA transcription factor that has a critical role in tissue differentiation and has been involved in tricho-rhino-phalangeal syndrome. It is highly expressed in some ER (-)

ductal breast carcinomas and contributes to epithelial to mesenchymal transitiicion in breast cancer (175).

-Mutations in the *ATP2A2* gene are associated with a class of keratosis follicularis (Darier disease), an autosomal dominant skin disorder characterized by loss of adhesion between epidermal cells and abnormal keratinization.

-*FAT1* encodes a member of the cadherin superfamily that acts as a tumor suppressor, essential for controlling cell proliferation in Drosophila development. It also acts as a tumor suppressor in humans, being involved in cell morphogenesis and migration. Mutations in this gene have been seen in oral squamous carcinoma (OSCC), affecting cell adhesion, migration and invasion (130)

-*ADAM10* encodes a transmembrane metalloproteinase that is involved in the proteolytic processing of NOTCH1 in skin epidermis(203). Therefore, it plays a role in the Notch signalling pathway. It has also been found to be upregulated in invading peripheral tumor cells of BCCs (133) and to be important in prostate cancer, where it can be accumulated in the nuclei acting as a transcription factor (7).

-*TNRC6B* plays a role in RNA-mediated gene silencing by both miRNAs and siRNAs. It is required for miRNA-dependent repression and siRNA dependent cleavage of complementary mRNAs by proteins of the argonaute family (106). At the moment, a role for this gene in cancer has not yet been established.

-*WAC* is involved in histone H2B ubiquitinization, and regulates transcriptional activation of the *CDKN1A* gene, playing a role in cell cycle checkpoint activation (217)

-The *SYNCRYP* gene encodes an hnRNP protein with RNA binding function involved in alternative splicing, polyadenilation and other aspects of mRNA metabolism and transport.

-*TP63* has a transcription factor function and plays an essential role in the development and maintenance of epithelial tissues. Mice lacking p63 present a profound block in the development of stratified epithelia and aplasia of multiple ectodermal appendages. A similar spectrum of phenotypic alterations is observed in human syndromes resulting from *TP63* gene mutations (151).

In summary, all this information supports the idea that these genes could have an important role in human skin carcinogenesis. Most of them genes have functions related to proliferation, cell morphogenesis, migration and cellular growth. So, mutation in these genes can lead to tumorigenesis. Based on this assumption, we decided to check the expression levels of these 11 genes in human tumor samples to identify a possible relationship between tumoral transformation and their expression levels.

6.1. qRT-PCR analysis

RNA was isolated from skin tumor samples and normal skin fragments adjacent to tumor areas. In total, we used 41 human skin tumors (30 BCCs and 11 SCCs.) and their paired control skins for analysis.

Results showed that *NOTCH*1 expression was downregulated in most of the human skin cancer samples, both in BCCs and SCCs (see **Figure 51**). These results strengthen the significance of the data that were obtained in the transposon insertion analysis in murine tumors, where Notch1 was the most frequently transposon-mutated gene. *ADAM10*, *TNRC6B* and *SYNCRYP* gene expression levels were frequently upregulated, while *TP63* gene expression was found to be downregulated both in BCC and SCC tumor samples (**Figure 51**).

We also found that some genes presented specific expression patterns associated to a certain type of tumour. For example, *NSD1* and *WAC* gene expression levels were mainly downregulated in BCCs and SCCs, respectively; *FAT1, ATP2A2* and *TRPS1* gene expression was upregulated primarily in BCCs. Other genes presented different gene expression alterations depending on tumor type. An example is *PARD3* which was often upregulated in BCCs but slightly downregulated in SCCs. This gene expression variability among tumor samples can be explained by intrinsic genetic differences inside each patient or within tumors, implying that not all tumors will have to be caused by alterations of the same pathways.

Statistical analysis indicated that only the relative expression values of *NSD1* and *GLI2* (used as a positive control that is amplified only in BCC) were significantly different between the BCC and SCC groups (t-test, p-value <0.05). These results could be attributed in part to the limited number of SCC samples available for the experiment and to the variability in the samples previously remarked.

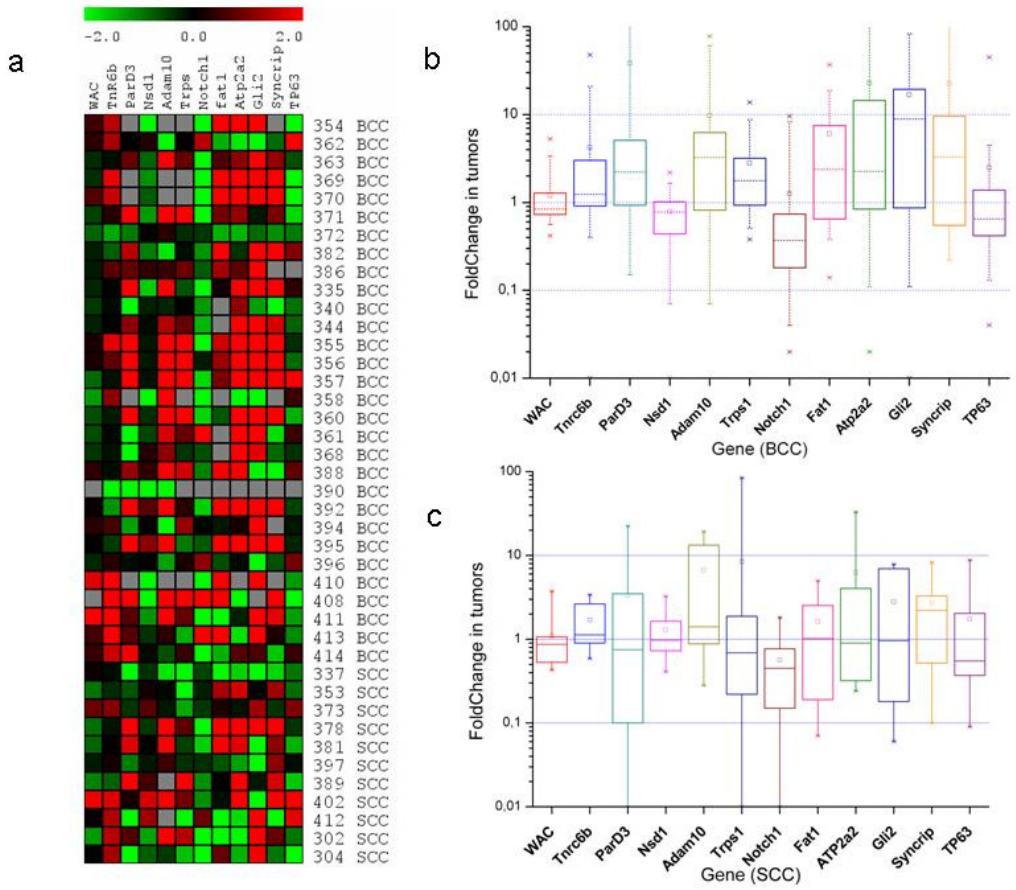

Figure 51. **qRT-PCR analysis of the expression of selected genes from CIS list and *GLI2* in human control skin and tumoral samples. a**) Heat map indicating the relative expression (tumor relative to adjacent healthy skin) of the analyzed genes; **b**) and **c**) Box and Whiskers plot of the relative expression values in BCCs (**b**) and SCCs (**c**)

6.2. Immunohistochemical analysis of NSD1 and PARD3 shows a reduced staining in BCC human samples compared to control skin

After proving that *NSD1* was clearly downregulated in a set of 30 human BCCs when compared to their respective normal skin samples, we decided to study its expression at the protein level in both BCCs and SCCs samples.

Therefore, we performed an immunohistochemical analysis to check the expression of NSD1 in a set of human skin control and tumor samples.

In control samples, NSD1 protein was detected in most of the nuclei of basal and suprabasal skin keratinocytes, (**Figure 52 a, b**). We found that in 100% of the assayed BCC tumor samples (N=13) NSD1 staining was greatly diminished or absent when compared to normal skin. This difference in staining levels was more evident in samples where control and tumor areas appeared together (**Figure 52, c, d**).

Immunohistochemical staining of SCCs samples revealed a diminished Nsd1 staining pattern in 63 % of tumor samples (7 out of 11 samples) when compared to normal skin tissue (**Figure 52e**) and also more noticeable when compared normal skin and tumor areas within the same sample (**Figure 52f).**

We also analyzed PARD3. Immunohistochemical staining of a total number of 7 BCCs samples using an antibody against PARD3 showed that cell membrane/cytoplasm PARD3 staining was decreased in 3 of the analyzed tumor samples (**Figure 52h**) compared to normal skin samples (**Figure 52g**), where PARD3 expression was present in the cell membrane/cytoplasm in most of the cell membrane/cytoplasm of basal and suprabasal skin cells.

In summary, the selected 16 genes present different expression levels among the human tumor samples. However, NSD1 expression is clearly downregulated in BCC and SCCS tumors and immunohistochemistry analysis show that NSD1 expression is very low or absent in BCCs. On the other side, PARD3 mRNA is upregulated in BCCs and slightly downregulated in SCCs, while its protein levels are lowered in some human BCCs. Futher studies are necessary involving a greater number of samples are necessary to characterize how these proteins are affect skin tumor development.

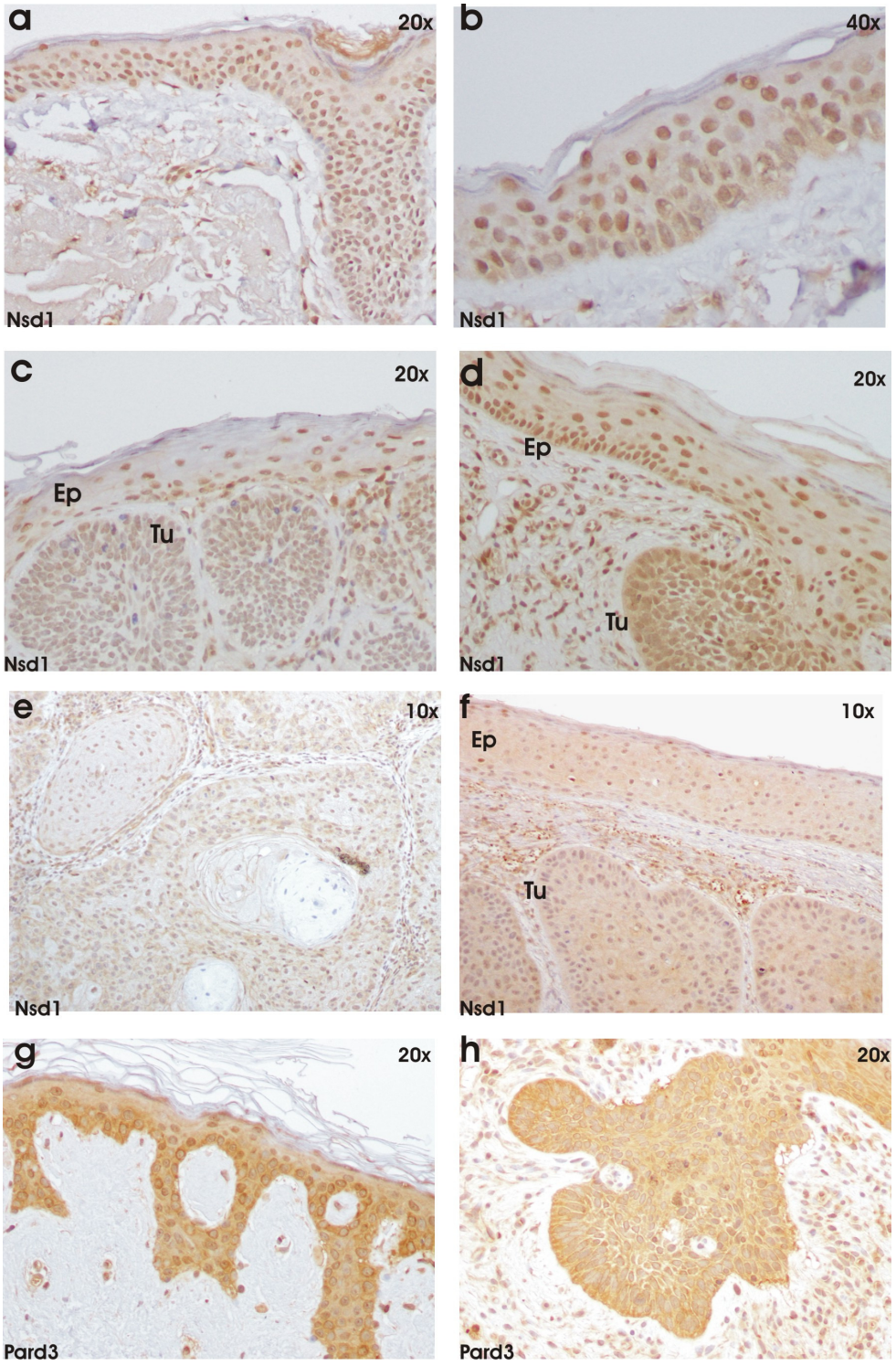

Figure 52. Immunohistochemistry analysis of NSD1 and PARD3 expression in control skin and human skin tumor samples. (**a**, **b**): NSD1 expression (brown) is seen in nuclei in normal skin samples; (**c**, **d**): Comparison of NSD1 expression levels in normal skin (**Ep**) and tumor areas (**Tu**) in BCC samples. NSD1 staining is generally diminished within the nuclei of tumor areas (**e** and **f**). PARD3 staining in cell membrane/cytoplasm is shown in normal control skin in **g**). In tumor samples (**h**) PARD3 staining appears to be reduced when compared to healthy skin samples (**g**)

7. ANALYSIS OF TUMOR DEVELOPMENT IN HETEROZYGOUS P53 MICE CONTAINING ACTIVE TRANSPOSONS

7.1. SB/T2/p53+/- mice developed more spontaneous mammary gland tumors than p53+/- mice without transposition

Since SB/T2 transgenic mice developed few spontaneous tumors, we decided to mate them with heterozygous p53+/- mice (FVB/J genetic background), prone to tumor development. We used both 6070 and 6113 transgenic lines of SB/T2 mice.

We found that SB/T2/p53+/- mice frequently suffered tumorigenesis in the mammary gland tissue. These mice also developed tumors of hematopoietic origin, something that has already been described in p53+/- mice (45), (86). Although the number of tumors from hematopoietic tissues could be considered high enough to be considered for analysis, we decide not to include them in our study because several transposon-based studies have been already done (32), (53), (18), while a transposon-based carcinogenesis analysis in mammary gland tissue has not yet been published.

Therefore, we monitorized mammary gland tumor development in SB/T2/p53+/- females and in control p53+/- females lacking transposition. We observed that from the first mating, the offspring of SB/T2/p53+/- animals developed mammary gland tumors with higher incidence than the control group formed by mice that were unable to undergo transposition (SB/p53+/-; T2/p53+/- and p53+/-). 41% of the SB/T2/p53+/- mice developed mammary gland tumors, while they were only found in 19.7% of the control animals. In addition, we saw that tumor development in SB/T2/p53+/- occurred after an average time of 49 weeks, while in control mice lacking transposition, this period increased up to 60 weeks. Statistically significant differences in mammary gland tumor-free survival were seen between experimental SB/T2/p53+/- and control group ($p<0.0002$). These results indicated that transposition facilitates the development of mammary gland tumors (**Figure 53**)

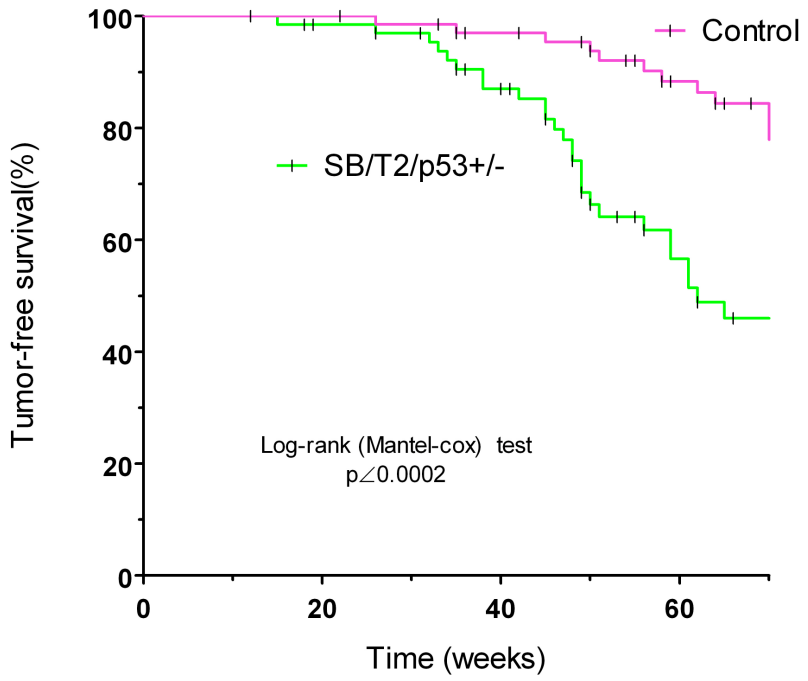

Figure 53. Kaplan-Meier analysis of mammary gland tumor free survival in SB/T2/p53+/- (n = 66) and in control p53+/- mice (n = 71). Animals that died without tumors, that died of non-mamary gland tumors or were alive at the end of the study were censored. Each censored animal is denoted by upward blips on the survival curves

Mammary gland tumors appeared only occasionally in SB/T2/p53+/+ mice background. In our study, we obtained a total number of 5 mammary gland tumors in this genetic background. These tumors were included with those obtained in SB/T2/p53+/- for further analysis, including histopathological characterization and analysis of transposon insertions within the tumors.

7.2. Histopathological characterization of the transposon/transposase mammary gland tumors revealed a common pattern of malignancy

We collected 38 mammary gland tumors from 33 female mice. The histopathological analysis of the collected tumors revealed that all were mammary gland carcinomas which were classified as solid, tubular, alveolar, acinar or adenosquamous based on their histological features (**Figure 54 and Figure 55**).

Most of the tumors presented a highly malignant profile with different levels of differentiation characterized by frequent mitosis, anisokaryosis and degeneration areas (**Figure 55**, **h, l, m**), metaplasia (**Figure 55m**), keratin pearl formation (**Figure 55f**) and also displayed signs of inflammation (**Figure 55l**) and neutrophil infiltration (**Figure 55k, asterisk**).

Histopathological classficiation		# Tumors
Carcinomas	Tubular	3
	Alveolar	9
	Tubular-alveolar	3
	Papilar	1
	Acinar	4
	Adenosquamous	4
	Solid	14
	Total	38

Figure 54. Histopathological classification of the 38 mammary gland tumors isolated from SB/T2/p53+/- and SB/T2 mice

7.3. SB11 expression in mammary gland tumors

After characterizing the mammary gland tumors pathology in SB/T2/p53+/- and SB/T2 mice, we decided to check SB11 expression in tumor samples from SB/T2/p53+/- mice by immunohistochemistry and compare it with K5 expression in the same samples. Since SB11 expression is controlled by regulatory sequences of K5 and K5 is only expressed in myoepithelial adult mammary gland cells, we wanted to see if the cell tumors were expressing the SB11 transposase.

Results showed SB11 expression in mammary gland myoephitelial cells in normal tissue (**Figure 56, a, b**) being consistent with the endogenous keratin K5 expression in these cells, except for the location of the proteins, being K5 expressed in cytoplasm and SB11 in the nuclei of the cells. SB11 protein was also expressed in the nuclei of some tumor cells in SB/T2/p53+/- mice (**Figure 56, d, f**), presenting a similar expression pattern to that of keratin K5 when analyzed in the same samples (**Figure 56, c, e**). While K5 protein was expressed in mammary gland tumor tissues p53+/- control mice (**Figure 56g**), the SB11 antibody gave no nuclear staining in the same samples (**Figure 56h**), proving the specificity of the nuclear signals obtained in SB/T2/p53+/- mice. Both this result and the different incidence of mammary gland carcinogenesis previously described strongly suggest that SB transposition play a role in mammary gland tumorigenesis in SB/T2/p53+/- mice.

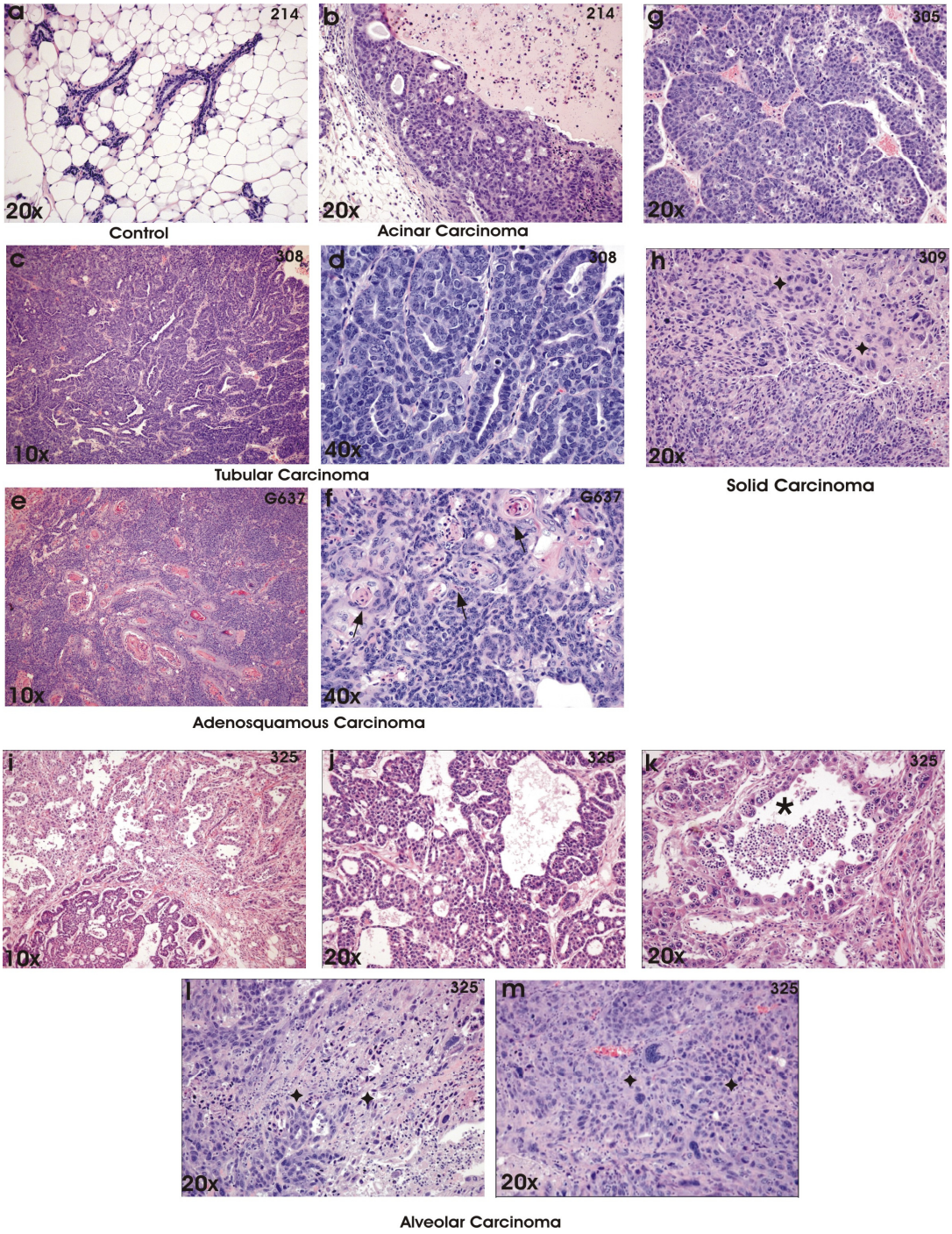

Figure 55. Hematoxolin and Eosin stained sections of a representative collection of mammary gland tumors showing the histopahological features of mammary gland tumors found in SB/T2/p53+/-(**b-d**) and (**g-m**); and SB/T2 mice (**e** and **f**). Stars mark regions of degeneration and anisokaryosis in (**h**, **l**, **m**). Asterisk marks region of inflammatory infiltrates in **k**. Arrows mark keratin pearl formations in **f** and degeneration areas in **m**

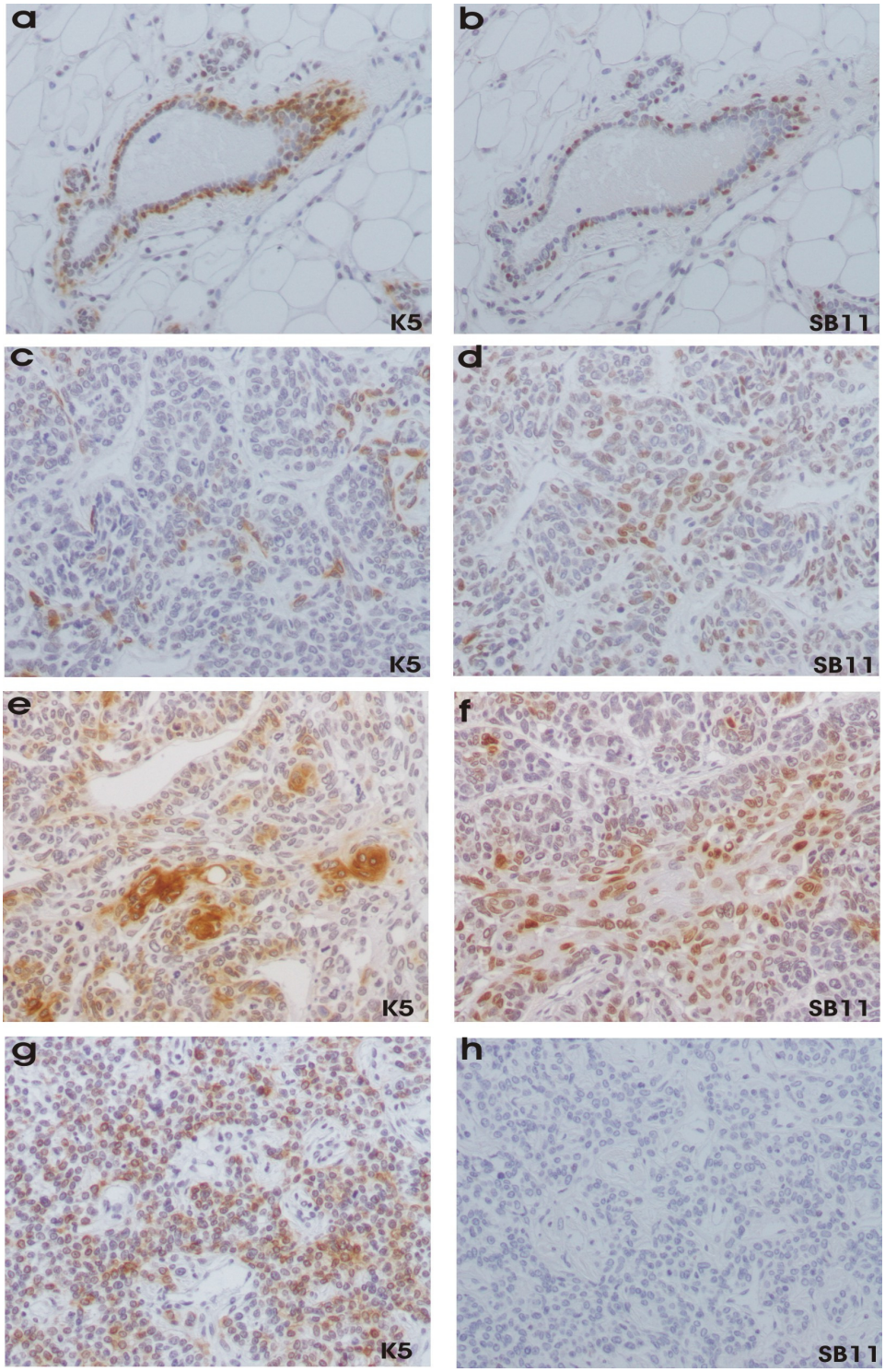

Figure 56. Immunohistochemistry analysis of K5 and SB11 expression in mammary gland healthy and tumor tissues. Histological sections from K5/SB/p53+/- mice (**a-f**) and p53+/- (**g, h**) were stained with K5 and SB11 antibodies (**brown staining**). All the photos correspond to a 20x microscope magnification

8. SEQUENCE ANALYSIS OF TRANSPOSON INSERTION SITES IN MAMMARY GLAND TUMORS

8.1. Amplification of the transposon-genomic junctions

Following the same experimental approach described for skin tumors, we next decided to analyze the transposon insertion sites in mammary tumors in order to identify genes mutated by transposon mobilization that may contribute to mammary gland carcinogenesis.

Genomic DNAs were isolated from the 38 mammary gland tumors isolated and were processed for high throughput sequencing as described in the Materials and Methods section. In order to check the quality of the generated genomic products before submitting the samples to sequencing, every PCR product was run on an agarose gel. Correct samples appeared as a smear of DNA with maximum intensity between 150bp and 500bp (**Figure 57**). After that, all the samples were pooled and sequence on a single run on a GAIIx Illumina sequencing platform.

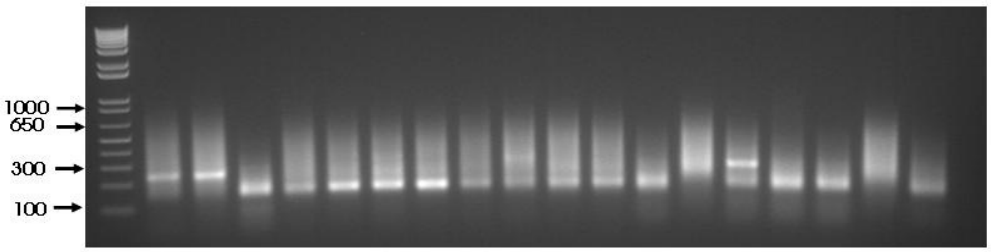

Figure 57. Agarose gel electrophoresis of a set of PCR amplified genomic fragments obtained from mammary gland tumor samples

8.2. Identification of CIS in the mammary gland tumors

For CIS identification in mammary gland tumors, we used the improved method called gene-centric CIS analysis (gCIS) that calculates the observed and expected number or insertion events within each RefSeq gene and assigns a p-value for each gene based on its statistical significance (27). For the analysis of CISs, we decided to divide the samples into two groups: one for tumors isolated from SB/T2 mice (n=5) and another one containing the tumors isolated from SB/T2/p53+/- mice (n = 33). This was done with the purpose of avoiding possible biases caused by the p53+/- background.

After performing the gCIS analysis we obtained a list of 51 CISs related to a specific gene region within the genome in the tumors from SB/T2/p53+/- mice group **(Figure 59)** and a list of 3 CISs from the tumors isolated from SB/T2 mice **(Figure 58).** An interesting result is that the

candidate cancer gene *Rasa1*, which was found at the top of both lists (Tables), is not dependent on the presence of a p53+/- background.

Another interesting result is that in both gCISs lists **(Figure 58 and 59)** most of the genes were predicted to be disrupted by the action of transposon insertions. This effect can be attributed to the preference of the transposon we used, T2Onc2, which has a promoter that is weakly active for oncogene activation, so it mainly causes inactivation of tumor suppressor genes.

Gene	# Tumors	p-value	Predicted Mutagenic effect
Map4k5	3	3.27E-79	Loss of function
Rasa1	3	1.60E-72	Loss of function
Trps1	3	6.20E-29	Loss of function

Figure 58. List of 3 gCIS gene list obtained from 5 tumors obtained from SB/T2 mice ordered by p-value significance

8.3. Candidate cancer genes in the gCIS list are enriched for signaling pathways and processes involved in carcinogenesis

After obtaining the gCIS list, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID) program (80) (79) to identify enriched biological processes associated with these candidate cancer genes.

The 51 genes were found to be associated with relevant Gene Ontology (GO) terms and statistically validated according to a *p- value* (0.01*)*, determining the probability that each biological function assigned to the dataset was caused by chance alone.

As it is represented in **(Figure 60),** organ development, gland development, transcription, gene expression and regulation of cell proliferation appeared as significantly enriched biological processes (BP) GO terms. Transcription factor, transcription activator, transcription regulator and nucleic acid binding were some of the most significantly enriched GO terms related to the molecular function (MF) category. And, in the cellular component (CC) category, the GO term nucleus appeared as one of the most relevant enriched terms.

These results suggest that these genes might play a role in carcinogenesis, since relevant enriched biological processes and functions, like organ development, gene expression, regulation of cell proliferation, transcription activation and nucleic acid binding are well known to be altered in tumorigenesis (71). Moreover, the Wnt signaling pathway was significantly enriched when looking at KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. This pathway has been related to tumorigenesis and specifically, to breast cancer (98), (77), (157), (122).

Gene	# Tumors	p-value	Predicted Mutagenic effect
Nf1	15	0	Loss of function
Rasa1	11	0	Loss of function
Nfib	7	1.39E-28	Loss of function
Axin1	6	9.58E-81	Loss of function
Ccnd3	5	2.12E-43	Gain of function
Pten	5	1.02E-37	Loss of function
Top1	5	1.69E-34	Loss of function
Znrf3	5	6.01E-20	Loss of function
Tead1	5	5.14E-18	Loss of function
Gata3	4	3.74E-88	Loss of function
Dyrk1a	4	6.48E-17	Loss of function
Ankhd1	4	1.04E-16	Loss of function
Mkln1	4	1.07E-15	Loss of function
Ankrd11	4	2.45E-15	Loss of function
Ankrd28	4	5.60E-13	Loss of function
Hydin	4	2.51E-07	Loss of function
4930412O13Rik	3	0	Loss of function
Krt6a	3	0	Loss of function
Rarg	3	1.34E-54	Gain of function
Ripk4	3	9.36E-46	Loss of function
Skint7	3	1.32E-37	Loss of function
Arf4	3	1.51E-33	Loss of function
Vangl1	3	2.44E-26	Loss of function
Acsl5	3	2.37E-25	Loss of function
Dsg1c	3	8.98E-23	Loss of function
Sfi1	3	1.02E-20	Gain of function
Stat5b	3	2.49E-20	Gain of function
Txnrd1	3	4.06E-20	Loss of function
Chst15	3	2.72E-19	Loss of function
Rnf43	3	1.17E-17	Loss of function
Lrp5	3	3.11E-17	Loss of function
Gatad2a	3	9.55E-17	Loss of function
Smad3	3	4.38E-16	Loss of function
Aff4	3	2.54E-14	Gain of function
Chd1	3	2.72E-14	Loss of function
Moxd1	3	7.80E-14	Loss of function
Zfr	3	1.17E-13	Loss of function
Rnf144a	3	2.52E-13	Gain of function
Snx27	3	3.13E-13	Gain of function
Setd2	3	1.69E-12	Loss of function
Flnb	3	5.87E-11	Loss of function
Usp6nl	3	7.10E-10	Loss of function
Ash1l	3	1.18E-09	Loss of function
Crebbp	3	2.98E-09	Loss of function
Cdh6	3	5.78E-09	Loss of function
Krt85	3	2.10E-08	Loss of function
Myo10	3	2.12E-08	Loss of function
Runx1	3	1.17E-07	Loss of function
Mef2a	3	1.78E-07	Loss of function
Fbxw7	3	1.15E-06	Loss of function
Foxn3	3	1.42E-06	Loss of function

Figure 59. List of 51 gCIS gene list from 33 tumors obtained from SB/T2/p53+/- mice ordered first by number of tumors that share a same CIS and then by p-value significance

TERM	BIOLOGICAL PROCESS	#GENES	*P*-VALUE
GO:0032502	Developmental process	25	1.11E-07
GO:0007275	Multicellular organismal development	23	6.12E-07
GO:0048513	Organ development	16	6.11E-05
GO:0045941	Positive regulation of transcription	9	7.57E-05
GO:0045944	Positive regulation of transcription from RNA polymerase II promoter	8	8.92E-05
GO:0010628	Positive regulation of gene expression	9	9.12E-05
GO:0048856	Anatomical structure development	18	9.77E-05
GO:0045935	Positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	9	1.24E-04
GO:0001953	Negative regulation of cell-matrix adhesion	3	1.33E-04
GO:0045893	Positive regulation of transcription, DNA-dependent	8	2.26E-04
GO:0009891	Positive regulation of biosynthetic process	9	2.27E-04
GO:0048523	Negative regulation of cellular process	12	6.42E-04
GO:0009893	Positive regulation of metabolic process	9	8.85E-04
GO:0048518	Positive regulation of biological process	13	1.00E-03
GO:0032501	Multicellular organismal process	24	1.00E-03
GO:0030155	Regulation of cell adhesion	4	2.00E-03
GO:0048732	Gland development	5	2.00E-03
GO:0008285	Negative regulation of cell proliferation	5	4.00E-03
GO:0006350	Transcription	13	5.00E-03
GO:0042127	Regulation of cell proliferation	7	5.00E-03
TERM	MOLECULAR FUNCTION	#GENES	*P*-VALUE
GO:0003700	transcription factor activity	11	3.36E-05
GO:0003677	DNA binding	16	3.90E-05
GO:0016563	transcription activator activity	7	6.73E-05
GO:0030528	transcription regulator activity	12	2.87E-04
GO:0005515	protein binding	27	0.001
GO:0005488	binding	40	0.002
GO:0003676	nucleic acid binding	17	0.003
TERM	CELLULAR COMPONENT	#GENES	*P*-VALUE
GO:0005634	nucleus	26	9.00E-05
GO:0043229	intracellular organelle	37	1.79E-04
GO:0043226	organelle	37	1.81E-04
GO:0005622	intracellular	41	3.33E-04
TERM	KEGG PATHWAYS	#GENES	*P*-VALUE
mmu04310	Wnt signaling pathway	6	7.20E-05
mmu05200	Pathways in cancer	6	0.002

Figure 60. Gene ontology (GO) terms significantly over-represented in the 51gCIS gene list. Only terms with p-value < 0.01 are included

9. VALIDATION OF CANDIDATE CANCER GENES BY IMMUNOHISTOCHEMISTRY IN HUMAN MAMMARY GLAND TISSUES

The final purpose of our study is to check if the candidate cancer genes that we have found in the SB/T2/p53+/- transgenic mouse model are relevant for human breast carcinogenesis. Although the list of gCIS is composed of 51 different genes, we have selected the two most representative genes at the top of the list for human validation because of several reasons. First, they were among the most frequently transposon-mutated genes in the gCIS list: *Nf1*, in 15 tumors and *Rasa1* in 11 tumors. Second, *Rasa1* gene was also mutated in tumors originated in SB/T2 mice, so we can assume transposon insertions in this gene were not biased by the p53+/- genetic background. And third, it has not yet.been established a clear relationship between these genes and breast tumorigenesis

9.1. *NF1* and *RASA1* as candidate cancer genes in human breast cancer

NF1 is a tumor suppressor gene that encodes the protein neurofibromin1, a negative regulator of the oncogenic Ras signal transduction pathway. Mutations in *NF1* have been linked to Neurofibromatosis type I, an autosomal genetic disorder with an estimated birth incidence of 1 in 2500(162).

According to COSMIC, *NF1* is somatically mutated mainly in soft tissues corresponding to neurofibromas and glomus tumors where 42.2% and 46.15%, respectively, of the analyzed samples were found to be mutated. *NF1* is mutated in a variety of tumor tissues like lung, ovary, intestine, hematopoietic, skin, prostate and breast, among others (see **Figure 61**).

Primary tissue	Unique mutated samples	Total unique samples	% Mutated
Breast	6	260	2.3
Liver	1	11	9.1
Large intestine	26	177	14.7
Central nervous system	47	701	6.7
Ovary	19	130	14.6
Haematopoietic and lymphoid tissue	20	326	6.1
Stomach	2	22	9.1
Soft_tissue	248	722	34.4
Kidney	2	206	1.0
Pancreas	1	48	2.1
Autonomic Ganglia	11	153	7.2
Skin	6	59	10.2
Prostate	1	63	1.6
Lung	25	548	4.6
Thyroid	1	11	9.1
Upper aerodigestive tract	4	26	15.4

Figure 61. Distribution of *NF1* somatic mutations in different tumor tissues acording to COSMIC database (v60). 42.4% of the soft_tissue samples were neurofibromas, 46.15% were glomus tumor and 22.4% were nerve tumors

Although most of the somatic mutations correspond to nervous system tissue, there are some studies that have found an increased risk of breast cancer in patients diagnosed with neurofibromatosis, (56, 162). Moreover, neurofibromin1 is absent in the human breast cancer line MDA-MP-231(132). Taken together, all these results strengthen the idea that *NF1* could have a role in mammary gland tumorigenesis.

RASA1 encodes the RAS GTPase activating protein p120GAP (also named as RASA1), a member of the GTP-ase activating proteins (GAPs) family. RASA1 protein activates the GTPase activity of wild type RAS protein, shifting it into its inactive GDP-bound form, being significantly important in cellular proliferation and considered to have a tumor suppressor activity. After looking at COSMIC database, *RASA1* appeared to be somatically mutated in a small number of tumors from skin, lung, ovary, prostate, upper aerodigestive tract and large intestine, being most of them classified as carcinomas. (**Figure 62**).

Primary tissue	Unique mutated samples	Total unique samples	% Mutated	Histology
Skin	1	7	14.3	Malignant melanoma
Ovary	1	59	1.7	Carcinoma
Prostate	2	61	3.3	Carcinoma
Lung	1	145	0.7	Carcinoma
Upper aerodigestive tract	5	7	71.4	Carcinoma
Large_intestine	5	17	29.4	Carcinoma

Figure 62. Distribution of *RASA1* somatic mutations in different tumor tissues according to COSMIC database (v60)

However, somatic mutations of this gene have also been observed in BCCs (62) and the gene is located in a region of frequent LOH in BRCA1 mutant breast cancer (89). This type of cancer has gene expression profiles similar to triple negative breast cancer subtype (which also harbour activating mutations of KRAS and BRAF). So mutations in *RASA1* could be associated to this breast cancer subtype.(78).

9.2. Study of the expression levels of NF1 and RASA1 in breast cancer human samples

As NF1 and RASA1 might have a role in human breast carcinogenesis, we decided to study the expression levels of NF1 and RASA1 proteins in human breast cancer and normal tissues by IHC, in order to check if their expression levels were reduced in human tumor samples

As it is shown in the **Figure 63,** staining of RASA1 protein in cytoplasm was lower in 100% of the analyzed tumor samples (4/4) **(Figure 63, b and c)** when compared to human control samples **(Figure 63a)**. Staining of NF1 protein in cytoplasm also appeared to be lower in 80% of the tumor samples (4/5) when compared to healthy samples **(Figure 63, e and d)**, being these staining level differences more marked than in the staining for RASA1. Therefore, these results suggest that *NF1* and *RASA1* genes might play a role in human breast carcinogenesis.

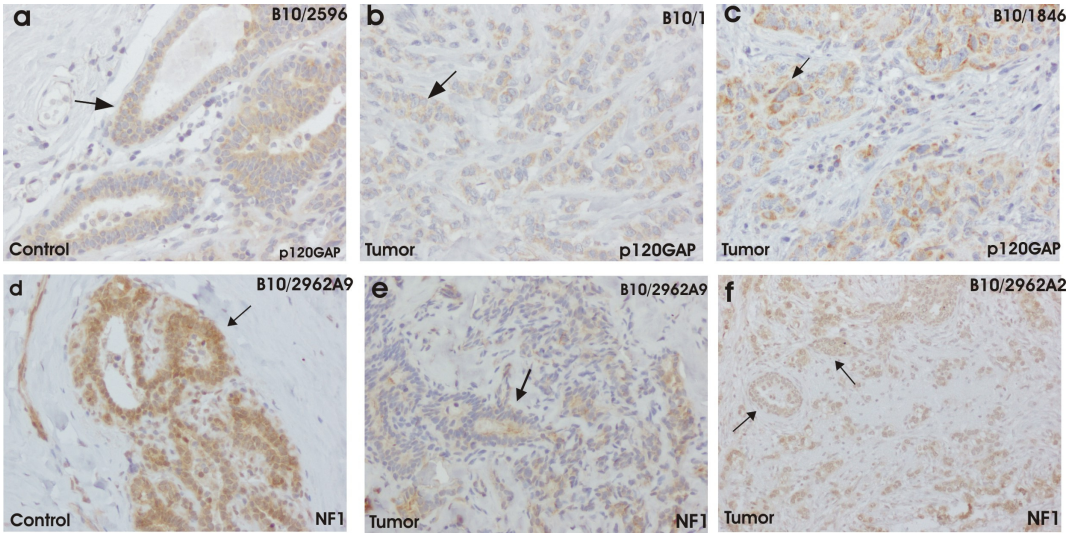

Figure 63. Immunohistochemical staining of human mammary gland tumors using antibodies that recognize NF1 and p120GAP proteins. **a** and **d** corresponds to healthy mammary gland tissue; **b**, **c**, **e** and **f** corresponds to tumorigenic areas from mammary gland carcinoma samples; d and e samples were isolated from the same patient. The arrows mark the NF1 and RASA1 cytoplasmic staining location

In summary, we have identified a list of candidate cancer genes that can be important for mammary gland tumorigenesis. Among them, we have characterized the expression levels of NF1 and RASA1 in human control and tumor samples and we have observed that expression levels are reduced in human tumor samples. These results correlate with the transposon insertions pattern obtained in mice mammary gland tumors, since loss of function was the predicted effect.

However, further studies are necessary in order to determine how a reduced expression of NF1 and RASA1 is related to human breast carcinogenesis.

DISCUSSION

1. TRANSPOSON MEDIATED MUTAGENESIS IN EPIDERMAL CELLS PROMOTES SKIN TUMORIGENESIS

In order to achieve transposon mobilization in the skin, we crossed transgenic mice carrying a SB11 transposase that was under the control of keratin K5 regulatory sequences to mice containing several copies of the T2/Onc2 mutagenic transposon.

Results from SB11 expression pattern in SB transgenic mice reveals that this protein is expressed in cells from the basal layer of the epidermis and from the outer root sheath and bulge of the follicles. Recent experiments have suggested that some skin cancer types could arise from epidermal stem cells residing in the basal layer and the hair follicle bulge. Our results proved that SB11 was expressed in both types of cells. Therefore, this model allows transposition mobilization and subsequent mutation in skin stem cells. In addition, SB11 expression was also found in other epithelial tissues expressing keratin K5, such as thymus, palate, ovary, lungs and mammary gland. These results suggest that transposition could affect not only to skin, but also to other organs where K5 is expressed, and in fact we have also found tumors in some of these tissues, analysis of which is beyond the scope of this work.

Our functional studies show that the SB transposition system works as expected. PCR and Southern Blot analysis (**Figure 35**) show that copies of the transposon T2Onc2 are mobilized by the SB11 transposase in the epidermis of the double transgenic SB/T2 mice. It is interesting that the transposition rate increases with time in the mice (**Figure 35**), considering that cancer develops through the sequential accumulation of mutations by normal cells. Therefore, this approach mimics this progression of the tumorigenesis process. However, only a few SB/T2 mice developed spontaneous skin tumors, even though we proved that transposition was happening in the SB/T2 mice epidermis. There are some reasons that could explain this result. First, the mice used in the experiment have a background (B6D2) that is not particularly sensitive to skin tumor development (72), (206). Second, previous experiments using the SB transposon system have shown that it is necessary a latency period to allow accumulation of transposition events in the cells for tumor formation in certain tissue types(33). And finally, the MSCV 5' LTR promoter used in the T2/Onc2, although active in keratinocites (11), is not specially strong. Subsequent experiments have proved that using a new transposon T2/Onc3, which contains the CAG promoter, which is expressed at high levels in epithelial cells, resulted in mice mostly developing a wide variety a carcinomas. However, this transposon was developed well after the beginning of our experiments, Therefore, in order to overcome these limitations, we performed the SB mutagenesis assay in mice of a Tg.AC genetic background,

more susceptible to skin tumor development or in B6D2 mice submitted to the well-established DMBA/TPA skin carcinogenesis protocol.

Analysis of papilloma development in mice after treatment with carcinogens showed that mice containing active transposons developed a more papillomas, with a shorter latency period and less regression than control animals lacking the SB transposase enzyme. In addition, histopathological analysis of the tumors showed that control animals developed the typical benign papillomas and occasional keratoacanthomas that are normally obtained using the DMBA/TPA carcinogenesis protocol. By contrast, most papillomas obtained from SB/T2 mice presented malignant alterations, many of which were located in basal cells, where the transposase SB11 is expressed. Moreover, SB/T2 mice developed BCCs and malignant SCCs. Interestingly, BCCs are never seen in wild type mice and never arise as a result of DBMA/TPA treatments. These results show that mobilization of the T2Onc2 transposon in the basal layer of the skin increases papilloma formation and malignancy and is able to generate BCCs.

Results obtained from Southern blot and sequencing analysis of a set of skin papillomas confirmed that new transposon insertions were clonally amplified in the tumors and that 75% of the sequenced insertion sites were located within genes in the genome. Several publications have proved that contrary to integrating retrovirus, the SB transposon does not present a significant preference for insertion near genes and instead, SB insertion sites are randomly distributed in the genome (113, 213). Based on this behavior, we can conclude that at least some of the transposon insertion sites in our mouse model are being positively selected during tumor formation because they confer a proliferative advantage to the tumor cells.

2. ANALYSIS OF TRANSPOSON INTEGRATION SITES IN THE SB/T2 SKIN TUMORS ALLOWS THE IDENTIFICATION OF NMSC CANDIDATE GENES

Pyrosequencing technology facilitates the use of transposons for cancer gene identification because it enables amplification and sequencing of tens of thousands of SB insertion sites from a mixture of tumors in a single sequencing run. In this work, we have identified a total number of 126 candidate cancer genes from a list of 108466 non redundant transposon insertion sites obtained from 75 skin tumors. Transposon insertion sites were also sequenced in the normal skin of two SB/T2 mice. Analysis of these insertions suggest that transposon mobilization is also happening in Wt skin but contrary to what happened in tumor samples, the number of insertions was very small and there was no positive selection over specific TA insertion sites.

The statistic tools to determine CIS regions from the transposon insertion sites mapped within the tumors assume that nondeleterous mutations will happen at random through the genome, but cancer driver mutations will cluster in these CIS regions that are characterized by having a higher mutation rate than expected by chance. We have obtained a total number of 126 CIS that encompass genes that might be important for skin tumorigenesis. Interestingly, no significant differences were seen in the genes mutated from Tg.AC TPA-treated and untreated or B6D2 TPA-treated mice, indicating probably that mouse background or TPA treatment do no require different mutations in each case to develop tumors. Additionally, most of the CIS regions identified in our study were found in less than 4 tumors. This large number of CIS shared by only a few tumors could mean that there are different alternative combinations of mutations that may lead to skin tumors, and can also be explained by different grades of papilloma progression involving the mutation of different genes and the intrinsic genetic tumor variability.

It is interesting to notice that our results present a low rate of local hopping transposition: 9.2% and 11% in T2/Onc2 lines 6070 and 6113 respectively, compared to what previous studies have obtained (95), (170), (171). This reduced local hopping rate can be attributed to the longer latency period of tumor development in the mice we used in our study. Since the transposase is expressing continuously, a long latency period for tumor development gives the transposon the chance to "hop" several times. Therefore, while many of the transposons will hop locally during the first rounds of transposition, many will hop to different chromosomes during the following rounds of transposition. In addition, the keratin K5 promoter drives high expression levels of SB11 transposase that could contribute to this effect. In fact, our analysis showed that double SB/T2 transgenic mice have mobilized at least once the majority of the copies of the transposon at the age of 8 weeks

Looking at the orientation of the mapped insertion sites, we also observed that most of the genes that we have identified in our study are predicted to be tumor suppressor genes. These results are understandable since we used a promoter that has not a strong activity in keratinocytes. Therefore, it probably results in slight activation of the oncogenes in which the transposon inserts, with the consequence that positively selected transposon insertions are predominantly those in "antisense" which cause gene inactivation. By contrast, recent experiments using a different transposon that has a CAG promoter (which presents a strong activity in epithelial tissues) identified mostly oncogenes, being *Zmiz1* the most frequently transposon-mutated gene (53). Interestingly, we have also identified *Zmiz1* in our study in few tumors. Consequently, our study complements the results described by Dupuy (53) by identifying mostly tumor suppressor genes.

Our data shows that the *Notch1* gene was by far the most frequently transposon-mutated locus. Three Notch paralogs (*Notch1*, *2* and *3*) are expressed in the epidermis and *Notch1* plays an important function in normal growth/differentiation control of keratinocytes (145). Numerous evidences link Notch1 to skin cancer. Reduction in Notch signaling within keratinocytes impairs their ability to execute the terminal differentiation program, resulting in formation of a defective skin-barrier and death if the areas involved are sufficiently large (21). It is been proposed that Notch1 mutation would prevent neoplastic cells from responding to differentiation signals in their immediate environment. Contrary to what happens in leukemia, where Notch1 acts as an oncogene, an opposite role for *Notch1* is observed in skin keratinocytes, where it acts as a tumor suppressor. Loss of *Notch1* in the epidermis causes development of mouse skin tumors that resemble BCCs and SCCs, in addition to inducing numerous papillomas (129), (139). It has been proprosed that deletion of *Notch1* in murine epidermis increases susceptibility to DMBA/TPA induced skin papillomas and that loss of Notch1 in keratinocytes infected with a retrovirus transducing the *Ras* oncogene and injected into nude mice can form aggressive SCC while wild type cells do not (129). In addition, it has been suggested that Notch1 mutation can substitute for TPA in the chemical carcinogenesis protocol (41). Most of the tumors used in our study were isolated from mice with Tg.AC background which harbor a mutated *H-ras* transgene that is transcriptionally silent in keratinocytes, but is activated upon proliferative stimuli, such as the TPA treatment. According to the proposed mechanism by which loss of *Notch1* activity may cause *Ras*-activation in skin cancer development and our data, mutation of *Notch1* could provide the stimuli necessary to activate *H-ras*, even if there is no TPA treatment. However, as said before, we did not observe significant differences in the genes mutated in tumors in both populations of SB/T2/Tg.AC mice, treated or not treated with TPA. In addition, most of the genes we have identified do not seem to be regulated in expression arrays from TPA-treated skin or TPA-generated tumors (40), (81).

It is interesting to mention that transposon mutations in the *Trp53* gene were much less frequent than many other genes in the 126CIS list and were only found in 2 tumors. Mutations in *P53* are frequent in human and murine skin tumors. However, most of these mutations are point mutations, in which a single nucleotide is substituted by another, and are located in “hotspot” areas, with a strong predominance in exons 5-8, which encode the DNA binding protein domain. The SB transposon mutagenesis mechanism is not well suited to produce point mutations. In addition, mutation of *Trp53* in mice treated with DMBA/TPA is not seen until later phases of the treatment, when benign papillomas convert to SCC tumors. Therefore, given that our system can not cause the “typical” p53 mutations, mutations in other genes that were important for cells to become tumorigenic were selected instead, such as mutations in *Notch1* or *Nsd1* genes.

In our results, we neither observed significant transposon mutations in the *PTCH1* gene, which is frequently mutated in human BCCs (63), (193), suggesting that there are alternative ways to develop this tumor in mice or perhaps there are other genes in this pathway whose mutation might also cause BCC. For instance, the *CSNK1A1* gene, an inhibitor of the *HH* pathway, is among the CIS genes identified in BCCs, as well as several members of the *Wnt* pathway, that has been postulated to interact with or cross-regulate the HH pathway (179), (131). In addition, it has also been proved that NOTCH1 deficiency in skin and primary keratinocytes causes Gli2 overexpression, leading to development of BCC-like tumors (129). Therefore, it is possible that the HH and Ras pathways interact in mouse skin (193).

The final purpose of this study is to find out if the genes we have found using a SB mutagenesis approach in mice are important for human cancer. Contrary to what happens in tumors from DMBA/TPA treated mice, *Ras* mutations are infrequent in human NMSC. However, mutations in the *NOTCH1* gene have been found in several forms of SCC (1), (176), (200). This was the most frequently transposon-mutated gene in our study and quantitative analysis of the *NOTCH1* gene expression in human BCCs and SCCs samples resulted in a strong down-regulation of the gene expression. Therefore, our study corroborates the importance of *NOTCH1* in human skin cancer development. In addition, several other genes that have been identified in this study encode proteins involved in *NOTCH1* regulation or function. For example, the *ADAM10* gene encodes a protein that acts as a crucial protease in the cleavage of Notch1. It has been proved that epidermal loss of *Adam10* results in impaired Notch processing in the epidermis and causes perinatal death and skin barrier defects in a knockout mouse model , pointing out an important role for *Adam10* in skin differentiation (203). The results of our study show that the *Adam10* gene was mutated by the transposon, and 100% of the insertions disrupted the CDS. However, the QPCR results indicate that this gene was overexpressed in human tumor samples. In a recent study it has been seen that ADAM10 protein is clearly upregulated in human BCCs (133), confirming our results in human tumor samples. It is possible that mutations affecting the *ADAM10* gene can create a non-functional version of the protein that will impair NOTCH1 processing, but is still recognized by antibodies in skin tumor samples. Considering the roles that NOTCH1 and ADAM10 have in the skin epidermis, it is possible that tumors in mice arose as a consequence of lack of control of differentiation, and other transposon-tagged genes might be also negative regulators of differentiation, allowing the precancerous keratinocytes to proliferate and expand their mutations through tumorigenesis process.

Another gene that was frequently transposon-mutated and has been related to NOTCH regulation is *Trp63*. It has been proved that a complex negative feedback loop exists between

NOTCH1 and TP63 that controls the balance between keratinoctye self-renewal and differentiation in human keratinocytes, being P63 expression suppressed when NOTCH1 is activated (128). Additional studies have found that *TP63* gene is overexpressed in a variety of epithelial tumors including oral and squamous carcinomas (204). However, our results of QPCR consistently showed a downregulation of this gene in BCC and SCC human samples. Given that our QPCR does not discriminate between the several TP63 isoforms and that we did not check the tumors by immunofluorescence, we can not discard a change in the TP63 subunit composition in the human skin tumors. In addition, it is known that *TP63* acts as a tumor suppressor in several tumors, such as prostate and bladder.

The second most transposon-mutated gene in mice skin tumors was *Nsd1,* which mainly underwent disruption by transposon insertions. Results from the QPCR analysis from human samples agreed with the results obtained in the transposition experiment and revealed that *NSD1* expression was clearly downregulated in BCC samples. In addition, IHC showed that NSD1 was absent or significantly reduced in 100% of the human BCCs samples analyzed, as compared to human control skin samples. In the case of human SCCs, 63% of samples presented reduced NSD1 expression levels. However, the QPCR analysis of NSD1 levels tested in SCCs samples did not present a clear downregulated expression pattern. This discrepancy can be attributed to the smaller number of SCC tumors used in the QPCR analysis and/or to the genetic intrinsic variability within the tumor samples. Nevertheless, these results suggest that reduced expression or loss of the *NSD1* gene might be important for NMSC development.

The *NSD1* gene encodes a hystone methyltransferase protein that mainly demethylates histone H3-lysine36 (H3K36) (112, 140), acting as an epigenetic regulator. Several studies that have linked this gene to carcinogenesis support the hypothesis of *NSD1* having a role in skin cancer development. First, Sotos patients, characterized by overgrowth features, present loss-of-function mutations (46) or haploinsufficiency in this gene (125) and have an increased risk for tumor development (181), including skin tumors(65). Second, chromosomic rearrarrangements of *NSD1* have been linked to several forms of cancer including the childhood acute myeloid leukemia where the NSD1-NUP98 fusion gene promotes *HOX* gene activation (87, 195), (198) and genomic alterations of *NSD1* have been also identified in breast cancer cell lines(220). And finally, several studies have linked the NSD1 functional status to different types of cancer: epigenetic inactivation of the *NSD1* promoter through CpG island hypermethylation is important in neuroblastomas and glioblastomas (17), and reduced mRNA and protein expression of *NSD1* gene correlated with a decrease in proliferation and augmented apoptosis in myeloma cell lines after treatment with an antiproliferative compound (219) and changes in the

expression of *NSD1* were found to be part of a "epigenetic gene signature" associated with prostate cancer progression(19).

However, it is not clear how *NSD1* loss of function can contribute to skin cancer tumorigenesis. Mice deficient in NSD1 exhibit an embryonic lethal phenotype with an enhanced apoptotic cell death in the ectoderm (146), implying that NSD1 is necessary for normal cell proliferation in mouse development. Additionally, it is known that the histone methyltransferase activity of NSD1 protein regulates transcriptional activation of adjacent target genes, such as *MEIS1* and *HOX* genes (198), (17). In a recent study, it has been shown that NSD1 binds to the promoter region of more than 300 genes in HTCC116 colon cancer cells, including known cancer genes and cell-cycle regulators, and that NSD1 depletion leads to a reduced expression of target genes (117). It is remarkable that among the NSD1 target genes identified in these colon cells, several were specific or related to skin, including keratin genes. Moreover, it has been seen that NSD1 physically interacts with ZMIZ4 and ZMIZ5 (202), two genes that belong to the same family than ZMIZ1, recently described to be involved in SCC tumor development using a SB-mutagenesis approach (53). Even though no epigenetic-related genes have yet been linked to the development of skin cancer, the *SETD2* gene, another H3K36 methyltransferase gene, is frequently mutated in renal carcinoma (39), (48). Supporting the idea that epigenetic modification could be causative in skin cancer, several of the transposon-tagged genes identified in this study have a role in epigenetics, such as the *WAC* gene, that binds to the RNA polymerase II complex regulating histone H2B ubiquitination, necessary for transcription(217). In addition, both *NSD1* and *WAC* genes were also identified as candidate cancer genes for colorectal cancer using the SB system (170), (171) and somatic mutations of NSD1 have been identified in head and neck SCC (176) and cutaneous SCC (54).

Pard3 was the third most mutated gene by transposon insertions in papillomas and the analysis of the transposon orientation pattern within the tumors revealed a preferential loss of function effect. However, results from the QPCR analysis showed that PARD3 expression was clearly upregulated in BCC tumors, while in SCC was slightly downregulated. But immunohistochemical results revealed that expression of PARD3 was diminished in more than half of the BCC tumor samples analyzed compared with normal skin samples. As it happened with *ADAM10* gene, it is possible that a non-functional protein is being overexpressed in human BCC tumors.

The *PARD3* gene encodes a protein expressed in the cytoplasm that interacts with other PAR family members affecting asymmetrical cell division and apical polarized cell growth. Epithelial cells have asymmetric distribution of cytoplasmic and membrane proteins to regulate cell

structure and transduce signals. Both tissue polarity and cell polarity are lost early during the neoplastic process due to alteration in polarity genes expression and/or their protein subcellular localization (109). Deletion of PARD3 in tumor cells may disrupt cell polarity and adhesion and some studies have supported this theory. Copy number loss of the *PARD3* gene was observed in 15% of primary esophageal squamous carcinoma cell lines and its expression was significantly reduced in tumors and correlated with aggressive tumor phenotypes (216). In another study, a significant low expression of PARD3 was seen in the distal margin of rectal cancer samples (116). Furthermore, inactivating mutations in both alleles of *PARD3* have been reported to occur in prostate cancer cells (96).

In addition, it has been proposed that alterations in the expression or functional activity of cell polarity proteins may prompt stem cells to divide symmetrically and to evade differentiation, providing a cancer stem cell-like environment (138). Stem cells undergo asymmetrical divisions by reorienting their mitotic spindles to generate a differentiated cell and in mammals, the dynamic interaction of PARD3 with a protein complex is key for the regulation of spindle orientation and cell differentiation (127). Therefore, loss in PARD3 expression could contribute to skin carcinogenesis by deregulation of asymmetrical cell division in the epidermal tissue. Moreover, in order to reorientate the mitotic spindle, cells respond to extracellular signals during which Notch signaling has a major role and it has been proved that reduction of the levels of some of these ligands result in the impairment of Notch signaling (190).
In conclusion, in this study we have identified several genes that are mutated in the transposon-generated tumor experiments and are presumably responsible for the development of these tumors. The analysis of the predicted role of these genes will allows us to infer that the processes are altered in skin tumors, including epigenetics.

3. TRANSPOSON MOBILIZATION PROMOTES MAMMARY GLAND TUMOR DEVELOPMENT IN HETEROZYGOUS P53 MICE

Although heterozygous p53 mice only develop a few mammary gland tumors, mostly due to an early death from sarcomas and lymphomas (45), it is known that *P53* is important for breast cancer development. A high proportion of patients with Li-Fraumeni syndrome carry germ-line mutations in the p53 gene that predispose them to a high risk of developing breast cancer, among others. Consistent with this information, it has been shown that approximately, 20% of all breast cancer cases present somatic mutations in p53 (155) and are most frequently found in the ductal carcinoma in situ (DCIS) stage (92).

Our results show that p53+/- mice carrying active transposons in keratin K5-positive cells developed mostly mammary gland carcinomas and these tumors arose earlier and with a statistically significant higher frequency than in control mice that did not undergo transposition,. In addition, we also found some mammary gland tumors in SB/T2/p53+/- animals. These results suggest that transposition in keratin k5-positive cells contribute to mammary gland tumor development.

In this study we have proved that the SB11 transposase is expressed in the myoephitelial cells of the adult mammary gland, following the same expression pattern of the keratin K5, except (as expected) for the nuclear localization of the SB11 protein. In addition, we have observed that SB11 is also expressed in mammary gland tumors from SB/T2/p53+/- mice. It is believed that most of the mammary gland tumors originate from mammary gland epithelial cells that in the adult mouse do not express K5. However, recent theories point out to the relevant contribution of mammary cancer stem cells in tumor development (3). This theory supports the idea that breast cancers are initiated and maintained by a subpopulation of tumor cells with stem cell features, known as cancer stem cells, characterized by self-renewal, differentiation and extensive capacity for proliferation properties. In the mouse, the existence of adult mammary gland stem cells was demonstrated by showing that a functional mammary gland could be generated in immunodeficient mice by transplantation of a single cell with stem cell like properties (161), (174). These mammary stem cells would be able to originate mammary gland cell lineages, luminal and myoepithelial cells and it is thought that these progenitor stem cells could be K5 positive (174), (221). Therefore, in our study, it is possible that transposition events took place in mammary gland stem cells that expressed K5 and consequently the SB11 transposase, originating tumors that are traditionally considered to arise from non-K5-expressing cells.

4. ANALISIS OF CIS IN THE SB/T2/P53+/- TUMORS LEADS TO IDENTIFICATION OF BREAST CANCER-CANDIDATE GENES

Recent advances in breast cancer genome sequencing have found driver mutations in 40 cancer genes and in 73 different combinations of mutated cancer genes among 100 breast tumors, highlighting the complexity of this genetic heterogeneous disease (173). Insertional mutagenesis using the SB transposon could allow distinguishing driving mutations from a large number of accompanying passenger mutations that do not contribute to cancer development.

In this study, after high throughput sequencing of 38 mammary gland tumors from mice undergoing active transposition we have obtained 51 candidate cancer genes. From the five SB/T2 mammary gland tumor samples we only identified 3 CIS. However, it is remarkable that one of the cancer gene candidates, *Map4k5*, showed up only in this mouse background, indicating that probably it participated in a pathway alternative to p53 for the development of cancer. The other two genes, *Trps1* and *Rasa1*, are present in both CIS lists from mice with and without p53+/- background, suggesting that the transposition insertions within these genes are not biased by the mouse genetic background (although Trps1 does not pass the statistical test in the group of p53+/- mice, it is also mutated in several tumors in this group). As it happened with the papilloma CIS list, based on the data of the orientation of the transposon with respect to the orientation of the mutated gene, most of the genes we identified were predicted to have a tumor suppressor function.

The results from the bioinformatics analysis on the DAVID platform revealed that the set of 51 candidate cancer genes was enriched in ontology categories related to transcriptional activation, regulation of cell proliferation and several cancer pathways. These data strongly support the idea of a positive selection in mammary gland tumors for transposon insertions in genes that are important for tumorigenesis. Among the enriched cancer pathways, it is interesting to mention the *Wnt* signaling pathway, which is known to be deregulated in breast cancer. The candidate cancer genes *Axin1*, *Lrp5*, *Ccnd3*, *Vangl1*, *Creebp* and *Smad3* were found to be significantly enriched for this pathway and *AXIN1* and *CCND3* were among the most mutated genes by transposon insertions.

Transgenic expression of b-catenin in the mammary gland leads to tumor formation(97). In human breast cancer the levels of this protein are increased and it has been shown that elevated transcriptional activity of the b-catenin gene contributes to the development of the disease (76). The AXIN1 protein functions as a negative regulator of the *Wnt* signaling pathway by being involved in the phosphorylation of b-catenin, which leads to ubiquitination and proteolysis of the protein. Therefore, a selection for transposon insertions that disrupt the *Axin1* gene with a loss-of-function effect will be a way to increase the levels of b-catenin that can lead to mammary gland tumorigenesis. Moreover, a recent work has related the loss of function of *AXIN1* to breast cancer since its lost caused overexpression of *MYC* oncogene in human breast cancer lines (14). The *Lrp5* gene also belongs to the *Wnt* signaling pathway and was also mutated by transposon insertions. This gene encodes a protein that acts as a correceptor with FRIZZLED protein to transduce signals by *Wnt* proteins, resulting in accumulation of an activated b-catenin. Recent studies have established a relationship between a non-functional LRP5 receptor and mammary gland tumor development (20). In addition, the VANGL1 gene encodes a planar cell polarity protein that acts as a *Wnt* signaling protein and recent experiments

have shown that increased levels of *VANGL1* transcripts correlates with risk of breast cancer relapse (5).

Although most of the genes were disrupted by transposon insertions with opposite orientation in relation to the targeted gene, *CCND3* gene was one exception, carrying transposon integrations in its same transcriptional orientation. This fact suggests that *CCND3* acts as an oncogene in mammary gland cancer. This gene encodes the Cyclin D3 protein, which is required for cell cycle G1/S transition. Our data agrees with the current scientific knowledge since this gene has been found to be overexpressed in human breast cancer cell lines (218), leading to an acceleration of cell cycle progression and subsequent increment in the cell proliferation rate that is associated with mammary gland tumor development.

Another of the most transposon-mutated genes was *Pten* and was found to be disrupted by transposon insertions. This tumor suppressor gene encodes a lipid phosphatase that dephosporylates PIP_3, antagonizing the PI3K/AKT pathway. Deletion of *PTEN* results in increased activation of the PI3K/AKT pathway that leads to elevated cellular levels of PIP_3 mimicking the effect of PI3K activation.and triggering the activation of its downstream effectors, PDK1, AKT/PKB and RAC1/CDC42. AKT activation stimulates cell cycle progression, survival, metabolism and migration (37), (47). Somatic mutations of *PTEN* are seen in only 6% of breast cancer samples (COSMIC), but LOH of *PTEN* is frequent in sporadic breast carcinoma (163) and 37% of breast tumors analyzed in a recent study presented no detectable or reduced levels of the protein (136). In addition, several mouse models have suggested that decreased *Pten* expression leads to an increased risk of breast tumor formation (111), (156). Therefore, our results are in agreement with the role of PTEN in mammary gland carcinogenesis and corroborate that in our study tumor cells selected CIS that were important for mammary gland tumor development.

The relevance of these candidate cancer genes identified in mammary gland tumors to human cancers was also tested by comparing the list of 53 genes with the list of human cancer genes contained in the Cancer Gene Census catalogue (http://www.sanger.ac.uk/genetics/CGP/Census). The results showed that 11 of the genes (21%) had a human orthologue in the catalogue (*NF1, NFIB, CCND3, PTEN, TOP1, GATA3, ANKHD1*, *ANKRD11, SETD2, CREBBP* and *FBXW7*). Therefore, they are recognized as human cancer genes. Furthermore, an analysis of the 53 genes using the COSMIC database (COSMIC v60) revealed that 52% of the genes were somatically mutated in breast carcinoma tumor samples and 4 of them were among the Top 20 most mutated genes (*PTEN*, *GATA3*, *CHD1* and *SETD2*), supporting the relevance of these genes for human breast cancer. In addition, several genes among the list of Cancer Gene Census that were not previously related to human breast cancer, have been identified to carry somatic driver mutations in a recent high throughput

sequencing study that involved 100 tumor samples (173), including *NF1*. Moreover, in a similar high throughput sequencing study involving more than 50 samples, another transposon-tagged gene (*RUNX1*), not previously observed in clinical breast cancer samples, has been recently found to be significantly mutated in breast cancer samples (13) and also correlated with oestrogen-receptor positive cancers (57).

We decided to focus our attention on the *NF1* and *RASA*1 genes for further functional studies and validation in human breast cancer samples for several reasons. First, *Nf1* and *Rasa1* were the most mutated genes by transposon insertions in the 51 gCIS list. Second, *Rasa1* has not been previously correlated to breast cancer, and although *Nf1*loss has been recently identified in an oncogenomis approach in mammary gland (196), when these experiments were performed that was not known. And finally, the *Rasa1* candidate cancer gene was identified as a CIS regardless of the mouse genetic background.

Our IHC results show that NF1 and RASA1 protein levels are significantly decreased in tumor samples compared to normal breast tissue protein immunostaining and consequently, these results agree with the positive selection of transposon insertions that disrupted the CDS of these genes in mammary gland tumors of SB/T2/p53+/- mice. Therefore, a loss of function of these predicted tumor suppressor genes could be important for human breast cancer development.

Even though a clear functional relationship between mutations in these genes and human breast cancer has not yet been established, some studies have pointed out relevant data that support this hypothesis. First, the Neurofibromatosis I disorder is linked to loss-of-function germline mutations in the *NF1* gene. This genetic autosomal disorder is characterized by the development of tumors from nervous tissue origin and some studies have reported an increase of breast cancer risk among neurofibromatosis female patients (162), (201). And second, as it has been mentioned previously, cancer high throughput sequencing studies have identified somatic intragenic mutations in this gene in clinical breast cancer samples (173).

The *NF1* gene encodes a GTPase protein that acts as a negative regulation of the RAS oncogene, driving it into its inactive GDP-bound form. Therefore, a non-functional NF1 protein could turn on constitutive activation of RAS, resulting in an increase of cell proliferation, characteristic of tumorigenesis. The *RASA1* gene encodes another GTPase protein that also acts as a suppressor of RAS function. Since activating *RAS* mutations are rare in human breast cancer, but the RAS protein is very often overexpressed in the tumors (55), our results suggest that transposon mutations in *Rasa1* and *Nf1* genes were selected because their disruption can lead to RAS constitutive activation and thus to an increase in cell proliferation rate. Moreover, results from the immunostaining analysis of NF1 and RASA1 proteins in human samples also indicate a reduced expression of these proteins in tumor samples and very recently, frequent mutations in NF1 have been identified in human breast carcinomas (196).

Further analysis to find out how *NF1*, *RASA1* and more candidate cancer genes identified by transposon insertions correlate with human breast tumor molecular profiles will be interesting, considering that breast cancer molecular subgroups have different tumor outcomes and prognosis. For example, the *NFIB* gene has been identified as a potential target for ER- human breast cancers (124). The *TEAD1* gene, that encodes for a transcription factor, has been correlated with poor clinical outcome in prostate cancer (100) and the*TRPS1* gene encodes a GATA transcription factor associated with ER- ductal carcinomas (175).

CONCLUSIONS

1. K5-SB11 transgenic mice express the SB11 transposase in basal cells of the epidermis and other stratified epithelia, in cells of the bulge region of the hair follicle (where some skin cancers are thought to originate from), and in myoepithelial cells of exocrine glands. So, this transgenic line can be used for mutagenesis studies in these cell types.

2. Double transgenic mice carrying K5-SB11 transgene and an integrated transgene array of T2/Onc2 mutagenic transposon (SB/T2 mice) undergo efficient transposon mobilization in the skin.

3. SB/T2 animals treated with TPA, either in wild-type or in H-ras mutated backgrounds, developed more malignant tumors of the skin and with decreased latency than control mice that did not undergo transposition. SB/T2 animals also developed basal cell carcinomas.

4. Analysis of the transposon insertion sites in the skin tumors identified 126 genes that were frequently mutated in different tumor samples and thus may have a role in human skin cancer development. Notch1 was the gene more frequently mutated.

5. Some of the most mutated genes in murine tumors also presented an altered expression pattern in human skin tumor samples. The expression level of NSD1 was decreased in most of the human skin tumor samples analyzed, compared to control samples.

6. SB/T2/p53+/- mice developed mammary gland tumors with higher incidence and with a shorter latency than mice that did not contain active transposons.

7. Analysis of transposon insertion sites in the mammary gland tumors from SB/T2/p53+/- mice identified 53 candidate genes that might be relevant for breast cancer development

8. Two of the genes of this list (*NF1* and *RASA1*) were found to be downregulated in human breast cancer samples, indicating that mutation in these genes could be important for the development of this pathology

CONCLUSIONES

1. Los ratones transgénicos K5-SB11 expresan transposasa en las células basales de la epidermis de la piel y otros epitelios estratificados, en las células de la vaina radicular externa, donde se cree que algunos tumores de piel se originan, y en células mioepiteliales de glándulas exocrinas. Por tanto, estos animales pueden ser usados para estudios de mutagénesis en estos tipos celulares.

2. En ratones dobles transgénicos portadores del transgén K5-SB11 y el concatémero de transposones mutagénicos T2/Onc2 (ratones SB/T2), se da una movilización de los transposones de manera eficaz.

3. Los animales SB/T2 sometidos a tratamientos de carcinogénesis de piel, tanto en fondo silvestre como portadores de un transgén H-ras activado, desarrollan tumores de piel más malignos con un menor período de latencia que animales sin transposición. Los animales SB/T2 también desarrollan carcinomas basocelulares.

4. El análisis de los sitios de integración del transposón en los tumores de piel identificó 126 genes mutados frecuentemente en diferentes tumores y que pueden por tanto tener un papel importante en el desarrollo de cáncer de piel. *Notch1* fue el gen más frecuentemente mutado.

5. Algunos de los genes más mutados en los tumores de ratón tenían un patrón de expresión alterado en muestras humanas de tumores de piel. En concreto, NSD1 se expresa a menor nivel en las muestras tumorales humanas que en muestras control.

6. Los ratones SB/T2/p53+/- desarrollaron tumores de glándula mamaria con mayor incidencia y con un período de latencia más corto que los ratones que no contenían transposones activos.

7. El análisis de los sitios de inserción de los transposones en los tumores de glándula mamaria de los ratones SB/T2p53+/- identificó 53 genes que pueden ser importantes para el desarrollo de cáncer de mama.

8. Dos de los genes de esta lista (*NF1* y *RASA1*) estaban reprimidos en muestras humanas de cáncer de mama, indicando que mutaciones en estos genes podrían ser importantes para el desarrollo de la patología.

1. **Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, Weinstein JN, Trevino L, Drummond JA, Muzny DM, Wu Y, Wood LD, Hruban RH, Westra WH, Koch WM, Califano JA, Gibbs RA, Sidransky D, Vogelstein B, Velculescu VE, Papadopoulos N, Wheeler DA, Kinzler KW, and Myers JN**. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science (New York, NY* 333: 1154-1157, 2011.
2. **Akagi K, Suzuki T, Stephens RM, Jenkins NA, and Copeland NG**. RTCGD: retroviral tagged cancer gene database. *Nucleic acids research* 32: D523-527, 2004.
3. **Al-Hajj M, and Clarke MF**. Self-renewal and solid tumor stem cells. *Oncogene* 23: 7274-7282, 2004.
4. **Aldaz CM, Trono D, Larcher F, Slaga TJ, and Conti CJ**. Sequential trisomization of chromosomes 6 and 7 in mouse skin premalignant lesions. *Molecular carcinogenesis* 2: 22-26, 1989.
5. **Anastas JN, Biechele TL, Robitaille M, Muster J, Allison KH, Angers S, and Moon RT**. A protein complex of SCRIB, NOS1AP and VANGL1 regulates cell polarity and migration, and is associated with breast cancer progression. *Oncogene* 31: 3696-3708, 2012.
6. **Andrechek ER, Hardy WR, Siegel PM, Rudnicki MA, Cardiff RD, and Muller WJ**. Amplification of the neu/erbB-2 oncogene in a mouse model of mammary tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America* 97: 3444-3449, 2000.
7. **Arima T, Enokida H, Kubo H, Kagara I, Matsuda R, Toki K, Nishimura H, Chiyomaru T, Tatarano S, Idesako T, Nishiyama K, and Nakagawa M**. Nuclear translocation of ADAM-10 contributes to the pathogenesis and progression of human prostate cancer. *Cancer science* 98: 1720-1726, 2007.
8. **Aszterbaum M, Epstein J, Oro A, Douglas V, LeBoit PE, Scott MP, and Epstein EH, Jr.** Ultraviolet and ionizing radiation enhance the growth of BCCs and trichoblastomas in patched heterozygous knockout mice. *Nature medicine* 5: 1285-1291, 1999.
9. **Aszterbaum M, Rothman A, Johnson RL, Fisher M, Xie J, Bonifas JM, Zhang X, Scott MP, and Epstein EH, Jr.** Identification of mutations in the human PATCHED gene in sporadic basal cell carcinomas and in patients with the basal cell nevus syndrome. *The Journal of investigative dermatology* 110: 885-888, 1998.
10. **Athar M, Tang X, Lee JL, Kopelovich L, and Kim AL**. Hedgehog signalling in skin development and cancer. *Experimental dermatology* 15: 667-677, 2006.
11. **Baldeschi C, Gache Y, Rattenholl A, Bouille P, Danos O, Ortonne JP, Bruckner-Tuderman L, and Meneguzzi G**. Genetic correction of canine dystrophic epidermolysis bullosa mediated by retroviral vectors. *Human molecular genetics* 12: 1897-1905, 2003.
12. **Balmain A, and Pragnell IB**. Mouse skin carcinomas induced in vivo by chemical carcinogens have a transforming Harvey-ras oncogene. *Nature* 303: 72-74, 1983.
13. **Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L, Cortes ML, Fernandez-Lopez JC, Peng S, Ardlie KG, Auclair D, Bautista-Pina V, Duke F, Francis J, Jung J, Maffuz-Aziz A, Onofrio RC, Parkin M, Pho NH, Quintanar-Jurado V, Ramos AH, Rebollar-Vega R, Rodriguez-Cuevas S, Romero-Cordoba SL, Schumacher SE, Stransky N, Thompson KM, Uribe-Figueroa L, Baselga J, Beroukhim R, Polyak K, Sgroi DC, Richardson AL, Jimenez-Sanchez G, Lander**

ES, Gabriel SB, Garraway LA, Golub TR, Melendez-Zajgla J, Toker A, Getz G, Hidalgo-Miranda A, and Meyerson M. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 486: 405-409, 2012.

14. **Bargonetti J, Friedman PN, Kern SE, Vogelstein B, and Prives C**. Wild-type but not mutant p53 immunopurified proteins bind to sequences adjacent to the SV40 origin of replication. *Cell* 65: 1083-1091, 1991.

15. **Bazopoulou D, and Tavernarakis N**. The NemaGENETAG initiative: large scale transposon insertion gene-tagging in Caenorhabditis elegans. *Genetica* 137: 39-46, 2009.

16. **Bender AM, Collier LS, Rodriguez FJ, Tieu C, Larson JD, Halder C, Mahlum E, Kollmeyer TM, Akagi K, Sarkar G, Largaespada DA, and Jenkins RB**. Sleeping beauty-mediated somatic mutagenesis implicates CSF1 in the formation of high-grade astrocytomas. *Cancer research* 70: 3557-3565, 2010.

17. **Berdasco M, Ropero S, Setien F, Fraga MF, Lapunzina P, Losson R, Alaminos M, Cheung NK, Rahman N, and Esteller M**. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proceedings of the National Academy of Sciences of the United States of America* 106: 21830-21835, 2009.

18. **Berquam-Vrieze KE, Nannapaneni K, Brett BT, Holmfeldt L, Ma J, Zagorodna O, Jenkins NA, Copeland NG, Meyerholz DK, Knudson CM, Mullighan CG, Scheetz TE, and Dupuy AJ**. Cell of origin strongly influences genetic selection in a mouse model of T-ALL. *Blood* 118: 4646-4656, 2011.

19. **Bianco-Miotto T, Chiam K, Buchanan G, Jindal S, Day TK, Thomas M, Pickering MA, O'Loughlin MA, Ryan NK, Raymond WA, Horvath LG, Kench JG, Stricker PD, Marshall VR, Sutherland RL, Henshall SM, Gerald WL, Scher HI, Risbridger GP, Clements JA, Butler LM, Tilley WD, Horsfall DJ, and Ricciardelli C**. Global levels of specific histone modifications and an epigenetic gene signature predict prostate cancer progression and development. *Cancer Epidemiol Biomarkers Prev* 19: 2611-2622, 2010.

20. **Bjorklund P, Svedlund J, Olsson AK, Akerstrom G, and Westin G**. The internally truncated LRP5 receptor presents a therapeutic target in breast cancer. *PloS one* 4: e4243, 2009.

21. **Blanpain C, Lowry WE, Pasolli HA, and Fuchs E**. Canonical notch signaling functions as a commitment switch in the epidermal lineage. *Genes & development* 20: 3022-3035, 2006.

22. **Bolos V, Peinado H, Perez-Moreno MA, Fraga MF, Esteller M, and Cano A**. The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. *Journal of cell science* 116: 499-511, 2003.

23. **Bolshakov S, Walker CM, Strom SS, Selvan MS, Clayman GL, El-Naggar A, Lippman SM, Kripke ML, and Ananthaswamy HN**. p53 mutations in human aggressive and nonaggressive basal and squamous cell carcinomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* 9: 228-234, 2003.

24. **Boukamp P**. Non-melanoma skin cancer: what drives tumor development and progression? *Carcinogenesis* 26: 1657-1667, 2005.

25. **Bray F, Ren JS, Masuyer E, and Ferlay J**. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *International journal of cancer* 2012.

26. **Bremner R, and Balmain A**. Genetic changes in skin tumor progression: correlation between presence of a mutant ras gene and loss of heterozygosity on mouse chromosome 7. *Cell* 61: 407-417, 1990.
27. **Brett BT, Berquam-Vrieze KE, Nannapaneni K, Huang J, Scheetz TE, and Dupuy AJ**. Novel molecular and computational methods improve the accuracy of insertion site analysis in Sleeping Beauty-induced tumors. *PloS one* 6: e24668, 2011.
28. **Campbell C, Quinn AG, Angus B, and Rees JL**. The relation between p53 mutation and p53 immunostaining in non-melanoma skin cancer. *The British journal of dermatology* 129: 235-241, 1993.
29. **Campbell C, Quinn AG, and Rees JL**. Codon 12 Harvey-ras mutations are rare events in non-melanoma human skin cancer. *The British journal of dermatology* 128: 111-114, 1993.
30. **Cano A, Perez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, Portillo F, and Nieto MA**. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature cell biology* 2: 76-83, 2000.
31. **Cardiff RD, Anver MR, Gusterson BA, Hennighausen L, Jensen RA, Merino MJ, Rehm S, Russo J, Tavassoli FA, Wakefield LM, Ward JM, and Green JE**. The mammary pathology of genetically engineered mice: the consensus report and recommendations from the Annapolis meeting. *Oncogene* 19: 968-988, 2000.
32. **Collier LS, Adams DJ, Hackett CS, Bendzick LE, Akagi K, Davies MN, Diers MD, Rodriguez FJ, Bender AM, Tieu C, Matise I, Dupuy AJ, Copeland NG, Jenkins NA, Hodgson JG, Weiss WA, Jenkins RB, and Largaespada DA**. Whole-body sleeping beauty mutagenesis can cause penetrant leukemia/lymphoma and rare high-grade glioma without associated embryonic lethality. *Cancer research* 69: 8429-8437, 2009.
33. **Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, and Largaespada DA**. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* 436: 272-276, 2005.
34. **Conti CJ, Aldaz CM, O'Connell J, Klein-Szanto AJ, and Slaga TJ**. Aneuploidy, an early event in mouse skin tumor development. *Carcinogenesis* 7: 1845-1848, 1986.
35. **Cooley L, Kelley R, and Spradling A**. Insertional mutagenesis of the Drosophila genome with single P elements. *Science (New York, NY* 239: 1121-1128, 1988.
36. **Crighton D, and Ryan KM**. Splicing DNA-damage responses to tumour cell death. *Biochimica et biophysica acta* 1705: 3-15, 2004.
37. **Dahia PL**. PTEN, a unique tumor suppressor gene. *Endocrine-related cancer* 7: 115-129, 2000.
38. **Dajee M, Lazarov M, Zhang JY, Cai T, Green CL, Russell AJ, Marinkovich MP, Tao S, Lin Q, Kubo Y, and Khavari PA**. NF-kappaB blockade and oncogenic Ras trigger invasive human epidermal neoplasia. *Nature* 421: 639-643, 2003.
39. **Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, Teague J, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Forbes S, Jia M, Jones D, Knott H, Kok CY, Lau KW, Leroy C, Lin ML, McBride DJ, Maddison M, Maguire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, O'Meara S, Pleasance E, Rajasingham A, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turrell K, Dykema KJ, Khoo SK, Petillo D, Wondergem B, Anema J, Kahnoski RJ, Teh BT,**

Stratton MR, and Futreal PA. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* 463: 360-363, 2010.

40. **Dang H, Trempus C, Malarkey DE, Wei SJ, Humble M, Morris RJ, and Tennant RW**. Identification of genes and gene ontology processes critical to skin papilloma development in Tg.AC transgenic mice. *Molecular carcinogenesis* 45: 126-140, 2006.

41. **Demehri S, Turkoz A, and Kopan R**. Epidermal Notch1 loss promotes skin tumorigenesis by impacting the stromal microenvironment. *Cancer cell* 16: 55-66, 2009.

42. **Depowski PL, Rosenthal SI, and Ross JS**. Loss of expression of the PTEN gene protein product is associated with poor outcome in breast cancer. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 14: 672-676, 2001.

43. **Derksen PW, Liu X, Saridin F, van der Gulden H, Zevenhoven J, Evers B, van Beijnum JR, Griffioen AW, Vink J, Krimpenfort P, Peterse JL, Cardiff RD, Berns A, and Jonkers J**. Somatic inactivation of E-cadherin and p53 in mice leads to metastatic lobular mammary carcinoma through induction of anoikis resistance and angiogenesis. *Cancer cell* 10: 437-449, 2006.

44. **DiGiovanni J**. Multistage carcinogenesis in mouse skin. *Pharmacology & therapeutics* 54: 63-128, 1992.

45. **Donehower LA, Harvey M, Slagle BL, McArthur MJ, Montgomery CA, Jr., Butel JS, and Bradley A**. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* 356: 215-221, 1992.

46. **Douglas J, Hanks S, Temple IK, Davies S, Murray A, Upadhyaya M, Tomkins S, Hughes HE, Cole TR, and Rahman N**. NSD1 mutations are the major cause of Sotos syndrome and occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes. *American journal of human genetics* 72: 132-143, 2003.

47. **Downward J**. PI 3-kinase, Akt and cell survival. *Seminars in cell & developmental biology* 15: 177-182, 2004.

48. **Duns G, van den Berg E, van Duivenbode I, Osinga J, Hollema H, Hofstra RM, and Kok K**. Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. *Cancer research* 70: 4287-4291, 2010.

49. **Dupuy AJ**. Transposon-based screens for cancer gene discovery in mouse models. *Semin Cancer Biol* 20: 261-268, 2010.

50. **Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, and Jenkins NA**. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* 436: 221-226, 2005.

51. **Dupuy AJ, Clark K, Carlson CM, Fritz S, Davidson AE, Markley KM, Finley K, Fletcher CF, Ekker SC, Hackett PB, Horn S, and Largaespada DA**. Mammalian germ-line transgenesis by transposition. *Proceedings of the National Academy of Sciences of the United States of America* 99: 4495-4499, 2002.

52. **Dupuy AJ, Fritz S, and Largaespada DA**. Transposition and gene disruption in the male germline of the mouse. *Genesis* 30: 82-88, 2001.

53. **Dupuy AJ, Rogers LM, Kim J, Nannapaneni K, Starr TK, Liu P, Largaespada DA, Scheetz TE, Jenkins NA, and Copeland NG**. A modified sleeping beauty transposon system that can be used to model a wide variety of human cancers in mice. *Cancer research* 69: 8150-8156, 2009.

54. **Durinck S, Ho C, Wang NJ, Liao W, Jakkula LR, Collisson EA, Pons J, Chan SW, Lam ET, Chu C, Park K, Hong SW, Hur JS, Huh N, Neuhaus IM, Yu SS, Grekin RC, Mauro TM, Cleaver JE, Kwok PY, Leboit PE, Getz G, Cibulskis**

K, Aster JC, Huang H, Purdom E, Li J, Bolund L, Arron ST, Gray JW, Spellman PT, and Cho RJ. Temporal Dissection of Tumorigenesis in Primary Cancers. *Cancer discovery* 1: 137-143, 2011.

55. **Eckert LB, Repasky GA, Ulku AS, McFall A, Zhou H, Sartor CI, and Der CJ**. Involvement of Ras activation in human breast cancer cell signaling, invasion, and anoikis. *Cancer research* 64: 4585-4592, 2004.

56. **el-Zawahry MD, Farid M, Abd el-Latif A, Horeia H, el-Gindy M, and Twakal G**. Breast lesions in generalized neurofibromatosis: breast cancer and cystosarcoma phylloides. *Neurofibromatosis* 2: 121-124, 1989.

57. **Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, Van Tine BA, Hoog J, Goiffon RJ, Goldstein TC, Ng S, Lin L, Crowder R, Snider J, Ballman K, Weber J, Chen K, Koboldt DC, Kandoth C, Schierding WS, McMichael JF, Miller CA, Lu C, Harris CC, McLellan MD, Wendl MC, DeSchryver K, Allred DC, Esserman L, Unzeitig G, Margenthaler J, Babiera GV, Marcom PK, Guenther JM, Leitch M, Hunt K, Olson J, Tao Y, Maher CA, Fulton LL, Fulton RS, Harrison M, Oberkfell B, Du F, Demeter R, Vickery TL, Elhammali A, Piwnica-Worms H, McDonald S, Watson M, Dooling DJ, Ota D, Chang LW, Bose R, Ley TJ, Piwnica-Worms D, Stuart JM, Wilson RK, and Mardis ER**. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486: 353-360, 2012.

58. **Erb P, Ji J, Kump E, Mielgo A, and Wernli M**. Apoptosis and pathogenesis of melanoma and nonmelanoma skin cancer. *Adv Exp Med Biol* 624: 283-295, 2008.

59. **Fickie MR, Lapunzina P, Gentile JK, Tolkoff-Rubin N, Kroshinsky D, Galan E, Gean E, Martorell L, Romanelli V, Toral JF, and Lin AE**. Adults with Sotos syndrome: review of 21 adults with molecularly confirmed NSD1 alterations, including a detailed case report of the oldest person. *American journal of medical genetics* 155A: 2105-2111, 2011.

60. **Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struewing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BA, Gayther SA, Zelada-Hedman M, and et al.** Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *American journal of human genetics* 62: 676-689, 1998.

61. **Friedenson B**. The BRCA1/2 pathway prevents hematologic cancers in addition to breast and ovarian cancers. *BMC cancer* 7: 152, 2007.

62. **Friedman E, Gejman PV, Martin GA, and McCormick F**. Nonsense mutations in the C-terminal SH2 region of the GTPase activating protein (GAP) gene in human tumours. *Nature genetics* 5: 242-247, 1993.

63. **Gailani MR, Stahle-Backdahl M, Leffell DJ, Glynn M, Zaphiropoulos PG, Pressman C, Unden AB, Dean M, Brash DE, Bale AE, and Toftgard R**. The role of the human homologue of Drosophila patched in sporadic basal cell carcinomas. *Nature genetics* 14: 78-81, 1996.

64. **Geurts AM, Yang Y, Clark KJ, Liu G, Cui Z, Dupuy AJ, Bell JB, Largaespada DA, and Hackett PB**. Gene transfer into genomes of human cells by the sleeping beauty transposon system. *Mol Ther* 8: 108-117, 2003.

65. **Gilaberte Y, Ferrer-Lozano M, Olivan MJ, Coscojuela C, Abascal M, and Lapunzina P**. Multiple giant pilomatricoma in familial Sotos syndrome. *Pediatric dermatology* 25: 122-125, 2008.

66. **Grachtchouk M, Mo R, Yu S, Zhang X, Sasaki H, Hui CC, and Dlugosz AA**. Basal cell carcinomas in mice overexpressing Gli2 in skin. *Nature genetics* 24: 216-217, 2000.
67. **Greenhalgh DA, Wang XJ, Donehower LA, and Roop DR**. Paradoxical tumor inhibitory effect of p53 loss in transgenic mice expressing epidermal-targeted v-rasHa, v-fos, or human transforming growth factor alpha. *Cancer research* 56: 4413-4423, 1996.
68. **Guy CT, Cardiff RD, and Muller WJ**. Activated neu induces rapid tumor progression. *The Journal of biological chemistry* 271: 7673-7678, 1996.
69. **Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, and Muller WJ**. Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proceedings of the National Academy of Sciences of the United States of America* 89: 10578-10582, 1992.
70. **Hahn H, Wicking C, Zaphiropoulous PG, Gailani MR, Shanley S, Chidambaram A, Vorechovsky I, Holmberg E, Unden AB, Gillies S, Negus K, Smyth I, Pressman C, Leffell DJ, Gerrard B, Goldstein AM, Dean M, Toftgard R, Chenevix-Trench G, Wainwright B, and Bale AE**. Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell* 85: 841-851, 1996.
71. **Hanahan D, and Weinberg RA**. The hallmarks of cancer. *Cell* 100: 57-70, 2000.
72. **Hennings H, Glick AB, Greenhalgh DA, Morgan DL, Strickland JE, Tennenbaum T, and Yuspa SH**. Critical aspects of initiation, promotion, and progression in multistage epidermal carcinogenesis. *Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine (New York, NY* 202: 1-8, 1993.
73. **Hennings H, and Yuspa SH**. Two-stage tumor promotion in mouse skin: an alternative interpretation. *Journal of the National Cancer Institute* 74: 735-740, 1985.
74. **Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, and Perou CM**. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome biology* 8: R76, 2007.
75. **Hogan B, Beddington R, Costantini F, and Lacy E**. Manipulating the mouse embryo, a laboratory manual. *2nd edn Cold Spring Harbor Laboratory Press Cold Spring Harbor, New York* 1994.
76. **Hollstein M, Sidransky D, Vogelstein B, and Harris CC**. p53 mutations in human cancers. *Science (New York, NY* 253: 49-53, 1991.
77. **Howe LR, and Brown AM**. Wnt signaling and breast cancer. *Cancer biology & therapy* 3: 36-41, 2004.
78. **Hu X, Stern HM, Ge L, O'Brien C, Haydu L, Honchell CD, Haverty PM, Peters BA, Wu TD, Amler LC, Chant J, Stokoe D, Lackner MR, and Cavet G**. Genetic alterations and oncogenic pathways associated with breast cancer subtypes. *Mol Cancer Res* 7: 511-522, 2009.
79. **Huang da W, Sherman BT, and Lempicki RA**. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37: 1-13, 2009.

80. **Huang da W, Sherman BT, and Lempicki RA**. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4: 44-57, 2009.
81. **Hummerich L, Muller R, Hess J, Kokocinski F, Hahn M, Furstenberger G, Mauch C, Lichter P, and Angel P**. Identification of novel tumour-associated genes differentially expressed in the process of squamous cell cancer development. *Oncogene* 25: 111-121, 2006.
82. **Hutchinson JN, Jin J, Cardiff RD, Woodgett JR, and Muller WJ**. Activation of Akt-1 (PKB-alpha) can accelerate ErbB-2-mediated mammary tumorigenesis but suppresses tumor invasion. *Cancer research* 64: 3171-3178, 2004.
83. **Isakoff SJ, Engelman JA, Irie HY, Luo J, Brachmann SM, Pearline RV, Cantley LC, and Brugge JS**. Breast cancer-associated PIK3CA mutations are oncogenic in mammary epithelial cells. *Cancer research* 65: 10992-11000, 2005.
84. **Ivics Z, Hackett PB, Plasterk RH, and Izsvak Z**. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 91: 501-510, 1997.
85. **Izsvak Z, Ivics Z, and Plasterk RH**. Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *Journal of molecular biology* 302: 93-102, 2000.
86. **Jacks T, Remington L, Williams BO, Schmitt EM, Halachmi S, Bronson RT, and Weinberg RA**. Tumor spectrum analysis in p53-mutant mice. *Curr Biol* 4: 1-7, 1994.
87. **Jaju RJ, Fidler C, Haas OA, Strickson AJ, Watkins F, Clark K, Cross NC, Cheng JF, Aplan PD, Kearney L, Boultwood J, and Wainscoat JS**. A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. *Blood* 98: 1264-1267, 2001.
88. **Jerry DJ, Kittrell FS, Kuperwasser C, Laucirica R, Dickinson ES, Bonilla PJ, Butel JS, and Medina D**. A mammary-specific model demonstrates the role of the p53 tumor suppressor gene in tumor development. *Oncogene* 19: 1052-1058, 2000.
89. **Johannsdottir HK, Jonsson G, Johannesdottir G, Agnarsson BA, Eerola H, Arason A, Heikkila P, Egilsson V, Olsson H, Johannsson OT, Nevanlinna H, Borg A, and Barkardottir RB**. Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. *International journal of cancer* 119: 1052-1060, 2006.
90. **Johnson RL, Rothman AL, Xie J, Goodrich LV, Bare JW, Bonifas JM, Quinn AG, Myers RM, Cox DR, Epstein EH, Jr., and Scott MP**. Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science (New York, NY* 272: 1668-1671, 1996.
91. **Jonkers J, Meuwissen R, van der Gulden H, Peterse H, van der Valk M, and Berns A**. Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nature genetics* 29: 418-425, 2001.
92. **Kastan MB, Onyekwere O, Sidransky D, Vogelstein B, and Craig RW**. Participation of p53 protein in the cellular response to DNA damage. *Cancer research* 51: 6304-6311, 1991.
93. **Kemp CJ, Donehower LA, Bradley A, and Balmain A**. Reduction of p53 gene dosage does not increase initiation or promotion but enhances malignant progression of chemically induced skin tumors. *Cell* 74: 813-822, 1993.
94. **Kemp CJ, Fee F, and Balmain A**. Allelotype analysis of mouse skin tumors using polymorphic microsatellites: sequential genetic alterations on chromosomes 6, 7, and 11. *Cancer research* 53: 6022-6027, 1993.

95. **Keng VW, Villanueva A, Chiang DY, Dupuy AJ, Ryan BJ, Matise I, Silverstein KA, Sarver A, Starr TK, Akagi K, Tessarollo L, Collier LS, Powers S, Lowe SW, Jenkins NA, Copeland NG, Llovet JM, and Largaespada DA**. A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nature biotechnology* 27: 264-274, 2009.
96. **Kern SE, Kinzler KW, Baker SJ, Nigro JM, Rotter V, Levine AJ, Friedman P, Prives C, and Vogelstein B**. Mutant p53 proteins bind DNA abnormally in vitro. *Oncogene* 6: 131-136, 1991.
97. **Kern SE, Kinzler KW, Bruskin A, Jarosz D, Friedman P, Prives C, and Vogelstein B**. Identification of p53 as a sequence-specific DNA-binding protein. *Science (New York, NY* 252: 1708-1711, 1991.
98. **Klaus A, and Birchmeier W**. Wnt signalling and its impact on development and cancer. *Nature reviews* 8: 387-398, 2008.
99. **Klein-Szanto AJ**. Pathology of human and experimental skin tumors. *Carcinogenesis; a comprehensive survey* 11: 19-53, 1989.
100. **Knight JF, Shepherd CJ, Rizzo S, Brewer D, Jhavar S, Dodson AR, Cooper CS, Eeles R, Falconer A, Kovacs G, Garrett MD, Norman AR, Shipley J, and Hudson DL**. TEAD1 and c-Cbl are novel prostate basal cell markers that correlate with poor clinical outcome in prostate cancer. *British journal of cancer* 99: 1849-1858, 2008.
101. **Kong J, Zhu F, Stalker J, and Adams DJ**. iMapper: a web application for the automated analysis and mapping of insertional mutagenesis sequence data against Ensembl genomes. *Bioinformatics* 24: 2923-2925, 2008.
102. **Kordon EC, and Smith GH**. An entire functional mammary gland may comprise the progeny from a single cell. *Development (Cambridge, England)* 125: 1921-1930, 1998.
103. **Langmead B, Trapnell C, Pop M, and Salzberg SL**. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* 10: R25, 2009.
104. **Largaespada DA, and Collier LS**. Transposon-mediated mutagenesis in somatic cells: identification of transposon-genomic DNA junctions. *Methods Mol Biol* 435: 95-108, 2008.
105. **Lazarov M, Kubo Y, Cai T, Dajee M, Tarutani M, Lin Q, Fang M, Tao S, Green CL, and Khavari PA**. CDK4 coexpression with Ras generates malignant human epidermal tumorigenesis. *Nature medicine* 8: 1105-1114, 2002.
106. **Lazzaretti D, Tournier I, and Izaurralde E**. The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins. *RNA (New York, NY* 15: 1059-1066, 2009.
107. **Leder A, Kuo A, Cardiff RD, Sinn E, and Leder P**. v-Ha-ras transgene abrogates the initiation step in mouse skin tumorigenesis: effects of phorbol esters and retinoic acid. *Proceedings of the National Academy of Sciences of the United States of America* 87: 9178-9182, 1990.
108. **Leder A, Pattengale PK, Kuo A, Stewart TA, and Leder P**. Consequences of widespread deregulation of the c-myc gene in transgenic mice: multiple neoplasms and normal development. *Cell* 45: 485-495, 1986.
109. **Lee M, and Vasioukhin V**. Cell polarity and cancer--cell and tissue polarity as a non-canonical tumor suppressor. *Journal of cell science* 121: 1141-1150, 2008.
110. **Li C, Chi S, He N, Zhang X, Guicherit O, Wagner R, Tyring S, and Xie J**. IFNalpha induces Fas expression and apoptosis in hedgehog pathway activated BCC cells through inhibiting Ras-Erk signaling. *Oncogene* 23: 1608-1617, 2004.

111. **Li Y, Podsypanina K, Liu X, Crane A, Tan LK, Parsons R, and Varmus HE**. Deficiency of Pten accelerates mammary oncogenesis in MMTV-Wnt-1 transgenic mice. *BMC molecular biology* 2: 2, 2001.
112. **Li Y, Trojer P, Xu CF, Cheung P, Kuo A, Drury WJ, 3rd, Qiao Q, Neubert TA, Xu RM, Gozani O, and Reinberg D**. The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *The Journal of biological chemistry* 284: 34283-34295, 2009.
113. **Liang Q, Kong J, Stalker J, and Bradley A**. Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. *Genesis* 47: 404-408, 2009.
114. **Lieu FM, Yamanishi K, Konishi K, Kishimoto S, and Yasuno H**. Low incidence of Ha-ras oncogene mutations in human epidermal tumors. *Cancer letters* 59: 231-235, 1991.
115. **Lin SC, Lee KF, Nikitin AY, Hilsenbeck SG, Cardiff RD, Li A, Kang KW, Frank SA, Lee WH, and Lee EY**. Somatic mutation of p53 leads to estrogen receptor alpha-positive and -negative mouse mammary tumors with high frequency of metastasis. *Cancer research* 64: 3525-3532, 2004.
116. **Liu J, Zhang W, Liu J, Lu X, Long Y, Zhou Y, and Liu S**. Expressions of connexin and par-3 in the distal margin of rectal cancer after ultra-low anterior resection. *Journal of Huazhong University of Science and Technology Medical sciences = Hua zhong ke ji da xue xue bao Yi xue Ying De wen ban = Huazhong keji daxue xuebao* 29: 330-334, 2009.
117. **Lucio-Eterovic AK, Singh MM, Gardner JE, Veerappan CS, Rice JC, and Carpenter PB**. Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proceedings of the National Academy of Sciences of the United States of America* 107: 16952-16957, 2010.
118. **Luo G, Ivics Z, Izsvak Z, and Bradley A**. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* 95: 10769-10773, 1998.
119. **Macias H, and Hinck L**. Mammary Gland Development. *Wiley interdisciplinary reviews Developmental biology* 1: 533-557, 2012.
120. **Mancuso M, Pazzaglia S, Tanori M, Hahn H, Merola P, Rebessi S, Atkinson MJ, Di Majo V, Covelli V, and Saran A**. Basal cell carcinoma and its development: insights from radiation-induced tumors in Ptch1-deficient mice. *Cancer research* 64: 934-941, 2004.
121. **Mann KM, Ward JM, Yew CC, Kovochich A, Dawson DW, Black MA, Brett BT, Sheetz TE, Dupuy AJ, Australian Pancreatic Cancer Genome I, Chang DK, Biankin AV, Waddell N, Kassahn KS, Grimmond SM, Rust AG, Adams DJ, Jenkins NA, and Copeland NG**. Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. *Proceedings of the National Academy of Sciences of the United States of America* 109: 5934-5941, 2012.
122. **Matsuda Y, Schlange T, Oakeley EJ, Boulay A, and Hynes NE**. WNT signaling enhances breast cancer cell motility and blockade of the WNT pathway by sFRP1 suppresses MDA-MB-231 xenograft growth. *Breast Cancer Res* 11: R32, 2009.
123. **Mattison J, van der Weyden L, Hubbard T, and Adams DJ**. Cancer gene discovery in mouse and man. *Biochimica et biophysica acta* 1796: 140-161, 2009.
124. **Moon HG, Hwang KT, Kim JA, Kim HS, Lee MJ, Jung EM, Ko E, Han W, and Noh DY**. NFIB is a potential target for estrogen receptor-negative breast cancers. *Molecular oncology* 5: 538-544.

125. **Nagai T, Matsumoto N, Kurotaki N, Harada N, Niikawa N, Ogata T, Imaizumi K, Kurosawa K, Kondoh T, Ohashi H, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Yokoyama T, Uetake K, Sakazume S, Fukushima Y, and Naritomi K**. Sotos syndrome and haploinsufficiency of NSD1: clinical features of intragenic mutations and submicroscopic deletions. *Journal of medical genetics* 40: 285-289, 2003.

126. **Nass SJ, and Dickson RB**. Defining a role for c-Myc in breast tumorigenesis. *Breast cancer research and treatment* 44: 1-22, 1997.

127. **Nelen MR, van Staveren WC, Peeters EA, Hassel MB, Gorlin RJ, Hamm H, Lindboe CF, Fryns JP, Sijmons RH, Woods DG, Mariman EC, Padberg GW, and Kremer H**. Germline mutations in the PTEN/MMAC1 gene in patients with Cowden disease. *Human molecular genetics* 6: 1383-1387, 1997.

128. **Nguyen BC, Lefort K, Mandinova A, Antonini D, Devgan V, Della Gatta G, Koster MI, Zhang Z, Wang J, Tommasi di Vignano A, Kitajewski J, Chiorino G, Roop DR, Missero C, and Dotto GP**. Cross-regulation between Notch and p63 in keratinocyte commitment to differentiation. *Genes & development* 20: 1028-1042, 2006.

129. **Nicolas M, Wolfer A, Raj K, Kummer JA, Mill P, van Noort M, Hui CC, Clevers H, Dotto GP, and Radtke F**. Notch1 functions as a tumor suppressor in mouse skin. *Nature genetics* 33: 416-421, 2003.

130. **Nishikawa Y, Miyazaki T, Nakashiro K, Yamagata H, Isokane M, Goda H, Tanaka H, Oka R, and Hamakawa H**. Human FAT1 cadherin controls cell migration and invasion of oral squamous cell carcinoma through the localization of beta-catenin. *Oncology reports* 26: 587-592, 2011.

131. **Noubissi FK, Goswami S, Sanek NA, Kawakami K, Minamoto T, Moser A, Grinblat Y, and Spiegelman VS**. Wnt signaling stimulates transcriptional outcome of the Hedgehog pathway by stabilizing GLI1 mRNA. *Cancer research* 69: 8572-8578, 2009.

132. **Ogata H, Sato H, Takatsuka J, and De Luca LM**. Human breast cancer MDA-MB-231 cells fail to express the neurofibromin protein, lack its type I mRNA isoform and show accumulation of P-MAPK and activated Ras. *Cancer letters* 172: 159-164, 2001.

133. **Oh ST, Schramme A, Stark A, Tilgen W, Gutwein P, and Reichrath J**. The disintegrin-metalloproteinases ADAM 10, 12 and 17 are upregulated in invading peripheral tumor cells of basal cell carcinomas. *Journal of cutaneous pathology* 36: 395-401, 2009.

134. **Oro AE, Higgins KM, Hu Z, Bonifas JM, Epstein EH, Jr., and Scott MP**. Basal cell carcinomas in mice overexpressing sonic hedgehog. *Science (New York, NY* 276: 817-821, 1997.

135. **Pasca di Magliano M, and Hebrok M**. Hedgehog signalling in cancer formation and maintenance. *Nature reviews* 3: 903-911, 2003.

136. **Perez-Tenorio G, Alkhori L, Olsson B, Waltersson MA, Nordenskjold B, Rutqvist LE, Skoog L, and Stal O**. PIK3CA mutations and PTEN loss correlate with similar prognostic factors and are not mutually exclusive in breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 13: 3577-3584, 2007.

137. **Polyak K**. Breast cancer: origins and evolution. *The Journal of clinical investigation* 117: 3155-3163, 2007.

138. **Prosser J, Porter D, Coles C, Condie A, Thompson AM, Chetty U, Steel CM, and Evans HJ**. Constitutional p53 mutation in a non-Li-Fraumeni cancer family. *British journal of cancer* 65: 527-528, 1992.
139. **Proweller A, Tu L, Lepore JJ, Cheng L, Lu MM, Seykora J, Millar SE, Pear WS, and Parmacek MS**. Impaired notch signaling promotes de novo squamous cell carcinoma formation. *Cancer research* 66: 7438-7444, 2006.
140. **Qiao Q, Li Y, Chen Z, Wang M, Reinberg D, and Xu RM**. The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *The Journal of biological chemistry* 286: 8361-8368, 2011.
141. **Quintanilla M, Brown K, Ramsden M, and Balmain A**. Carcinogen-specific mutation and amplification of Ha-ras during mouse skin carcinogenesis. *Nature* 322: 78-80, 1986.
142. **Rahrmann EP, Collier LS, Knutson TP, Doyal ME, Kuslak SL, Green LE, Malinowski RL, Roethe L, Akagi K, Waknitz M, Huang W, Largaespada DA, and Marker PC**. Identification of PDE4D as a proliferation promoting factor in prostate cancer using a Sleeping Beauty transposon-based somatic mutagenesis screen. *Cancer research* 69: 4388-4397, 2009.
143. **Ramirez A, Bravo A, Jorcano JL, and Vidal M**. Sequences 5' of the bovine keratin 5 gene direct tissue- and cell-type-specific expression of a lacZ gene in the adult and during development. *Differentiation* 58: 53-64, 1994.
144. **Ramirez A, Page A, Gandarillas A, Zanet J, Pibre S, Vidal M, Tusell L, Genesca A, Whitaker DA, Melton DW, and Jorcano JL**. A keratin K5Cre transgenic line appropriate for tissue-specific or generalized Cre-mediated recombination. *Genesis* 39: 52-57, 2004.
145. **Rangarajan A, Talora C, Okuyama R, Nicolas M, Mammucari C, Oh H, Aster JC, Krishna S, Metzger D, Chambon P, Miele L, Aguet M, Radtke F, and Dotto GP**. Notch signaling is a direct determinant of keratinocyte growth arrest and entry into differentiation. *The EMBO journal* 20: 3427-3436, 2001.
146. **Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, Lerouge T, Hager GL, Chambon P, and Losson R**. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *The EMBO journal* 22: 3153-3163, 2003.
147. **Regl G, Kasper M, Schnidar H, Eichberger T, Neill GW, Philpott MP, Esterbauer H, Hauser-Kronberger C, Frischauf AM, and Aberger F**. Activation of the BCL2 promoter in response to Hedgehog/GLI signal transduction is predominantly mediated by GLI2. *Cancer research* 64: 7724-7731, 2004.
148. **Reifenberger J, Wolter M, Knobbe CB, Kohler B, Schonicke A, Scharwachter C, Kumar K, Blaschke B, Ruzicka T, and Reifenberger G**. Somatic mutations in the PTCH, SMOH, SUFUH and TP53 genes in sporadic basal cell carcinomas. *The British journal of dermatology* 152: 43-51, 2005.
149. **Reinisch CM, Uthman A, Erovic BM, and Pammer J**. Expression of BMI-1 in normal skin and inflammatory and neoplastic skin lesions. *Journal of cutaneous pathology* 34: 174-180, 2007.
150. **Riobo NA, Lu K, Ai X, Haines GM, and Emerson CP, Jr.** Phosphoinositide 3-kinase and Akt are essential for Sonic Hedgehog signaling. *Proceedings of the National Academy of Sciences of the United States of America* 103: 4505-4510, 2006.
151. **Romano RA, Solomon LW, and Sinha S**. Tp63 in oral development, neoplasia, and autoimmunity. *Journal of dental research* 91: 125-132, 2012.
152. **Rothenberg SM, Mohapatra G, Rivera MN, Winokur D, Greninger P, Nitta M, Sadow PM, Sooriyakumar G, Brannigan BW, Ulman MJ, Perera RM, Wang R,**

Tam A, Ma XJ, Erlander M, Sgroi DC, Rocco JW, Lingen MW, Cohen EE, Louis DN, Settleman J, and Haber DA. A genome-wide screen for microdeletions reveals disruption of polarity complex genes in diverse human cancers. *Cancer research* 70: 2158-2164, 2010.

153. **Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, and Quackenbush J**. TM4 microarray software suite. *Methods in enzymology* 411: 134-193, 2006.

154. **Saridaki Z, Liloglou T, Zafiropoulos A, Koumantaki E, Zoras O, and Spandidos DA**. Mutational analysis of CDKN2A genes in patients with squamous cell carcinoma of the skin. *The British journal of dermatology* 148: 638-648, 2003.

155. **Saylors RL, 3rd, Sidransky D, Friedman HS, Bigner SH, Bigner DD, Vogelstein B, and Brodeur GM**. Infrequent p53 gene mutations in medulloblastomas. *Cancer research* 51: 4721-4723, 1991.

156. **Schade B, Rao T, Dourdin N, Lesurf R, Hallett M, Cardiff RD, and Muller WJ**. PTEN deficiency in a luminal ErbB-2 mouse model results in dramatic acceleration of mammary tumorigenesis and metastasis. *The Journal of biological chemistry* 284: 19018-19026, 2009.

157. **Schlange T, Matsuda Y, Lienhard S, Huber A, and Hynes NE**. Autocrine WNT signaling contributes to breast cancer cell proliferation via the canonical WNT pathway and EGFR transactivation. *Breast Cancer Res* 9: R63, 2007.

158. **Schnidar H, Eberl M, Klingler S, Mangelberger D, Kasper M, Hauser-Kronberger C, Regl G, Kroismayr R, Moriggl R, Sibilia M, and Aberger F**. Epidermal growth factor receptor signaling synergizes with Hedgehog/GLI in oncogenic transformation via activation of the MEK/ERK/JUN pathway. *Cancer research* 69: 1284-1292, 2009.

159. **Schoenenberger CA, Andres AC, Groner B, van der Valk M, LeMeur M, and Gerlinger P**. Targeted c-myc gene expression in mammary glands of transgenic mice induces mammary tumours with constitutive milk protein gene transcription. *The EMBO journal* 7: 169-175, 1988.

160. **Seitz CS, Lin Q, Deng H, and Khavari PA**. Alterations in NF-kappaB function in transgenic epithelial tissue demonstrate a growth inhibitory role for NF-kappaB. *Proceedings of the National Academy of Sciences of the United States of America* 95: 2307-2312, 1998.

161. **Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, Wu L, Lindeman GJ, and Visvader JE**. Generation of a functional mammary gland from a single stem cell. *Nature* 439: 84-88, 2006.

162. **Sharif S, Moran A, Huson SM, Iddenden R, Shenton A, Howard E, and Evans DG**. Women with neurofibromatosis 1 are at a moderately increased risk of developing breast cancer and should be considered for early screening. *Journal of medical genetics* 44: 481-484, 2007.

163. **Singh B, Ittmann MM, and Krolewski JJ**. Sporadic breast cancers exhibit loss of heterozygosity on chromosome segment 10q23 close to the Cowden disease locus. *Genes, chromosomes & cancer* 21: 166-171, 1998.

164. **Sinn E, Muller W, Pattengale P, Tepler I, Wallace R, and Leder P**. Coexpression of MMTV/v-Ha-ras and MMTV/c-myc genes in transgenic mice: synergistic action of oncogenes in vivo. *Cell* 49: 465-475, 1987.

165. **Slaga TJ, O'Connell J, Rotstein J, Patskan G, Morris R, Aldaz CM, and Conti CJ**. Critical genetic determinants and molecular events in multistage skin carcinogenesis. *Symposium on Fundamental Cancer Research* 39: 31-44, 1986.

166. **Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A, and et al.** Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science (New York, NY* 244: 707-712, 1989.
167. **Southern EM**. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of molecular biology* 98: 503-517, 1975.
168. **Soyal SM, Mukherjee A, Lee KY, Li J, Li H, DeMayo FJ, and Lydon JP**. Cre-mediated recombination in cell lineages that express the progesterone receptor. *Genesis* 41: 58-66, 2005.
169. **Spalding JW, Momma J, Elwell MR, and Tennant RW**. Chemically induced skin carcinogenesis in a transgenic mouse line (TG.AC) carrying a v-Ha-ras gene. *Carcinogenesis* 14: 1335-1341, 1993.
170. **Starr TK, Allaei R, Silverstein KA, Staggs RA, Sarver AL, Bergemann TL, Gupta M, O'Sullivan MG, Matise I, Dupuy AJ, Collier LS, Powers S, Oberg AL, Asmann YW, Thibodeau SN, Tessarollo L, Copeland NG, Jenkins NA, Cormier RT, and Largaespada DA**. A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science (New York, NY* 323: 1747-1750, 2009.
171. **Starr TK, Scott PM, Marsh BM, Zhao L, Than BL, O'Sullivan MG, Sarver AL, Dupuy AJ, Largaespada DA, and Cormier RT**. A Sleeping Beauty transposon-mediated screen identifies murine susceptibility genes for adenomatous polyposis coli (Apc)-dependent intestinal tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America* 108: 5765-5770, 2011.
172. **Steeg PS, and Zhou Q**. Cyclins and breast cancer. *Breast cancer research and treatment* 52: 17-28, 1998.
173. **Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, Butler A, Cheverton A, Gamble J, Hinton J, Jia M, Jayakumar A, Jones D, Latimer C, Lau KW, McLaren S, McBride DJ, Menzies A, Mudie L, Raine K, Rad R, Chapman MS, Teague J, Easton D, Langerod A, Lee MT, Shen CY, Tee BT, Huimin BW, Broeks A, Vargas AC, Turashvili G, Martens J, Fatima A, Miron P, Chin SF, Thomas G, Boyault S, Mariani O, Lakhani SR, van de Vijver M, van 't Veer L, Foekens J, Desmedt C, Sotiriou C, Tutt A, Caldas C, Reis-Filho JS, Aparicio SA, Salomon AV, Borresen-Dale AL, Richardson AL, Campbell PJ, Futreal PA, and Stratton MR**. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486: 400-404, 2012.
174. **Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li HI, and Eaves CJ**. Purification and unique properties of mammary epithelial stem cells. *Nature* 439: 993-997, 2006.
175. **Stinson S, Lackner MR, Adai AT, Yu N, Kim HJ, O'Brien C, Spoerke J, Jhunjhunwala S, Boyd Z, Januario T, Newman RJ, Yue P, Bourgon R, Modrusan Z, Stern HM, Warming S, de Sauvage FJ, Amler L, Yeh RF, and Dornan D**. miR-221/222 targeting of trichorhinophalangeal 1 (TRPS1) promotes epithelial-to-mesenchymal transition in breast cancer. *Science signaling* 4: pt5, 2011.
176. **Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, Shefler E, Ramos AH, Stojanov P, Carter SL, Voet D, Cortes ML, Auclair D, Berger MF, Saksena G, Guiducci C, Onofrio RC, Parkin M, Romkes M, Weissfeld JL, Seethala RR, Wang L, Rangel-Escareno C, Fernandez-Lopez JC, Hidalgo-Miranda A, Melendez-Zajgla J, Winckler W, Ardlie K, Gabriel SB, Meyerson M, Lander ES, Getz G,**

Golub TR, Garraway LA, and Grandis JR. The mutational landscape of head and neck squamous cell carcinoma. *Science (New York, NY* 333: 1157-1160, 2011.

177. **Sun P, Yuan Y, Li A, Li B, and Dai X**. Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. *Histochemistry and cell biology* 133: 213-221, 2010.

178. **Sutherland RL, and Musgrove EA**. Cyclin D1 and mammary carcinoma: new insights from transgenic mouse models. *Breast Cancer Res* 4: 14-17, 2002.

179. **Taipale J, and Beachy PA**. The Hedgehog and Wnt signalling pathways in cancer. *Nature* 411: 349-354, 2001.

180. **Tang JY, Aszterbaum M, Athar M, Barsanti F, Cappola C, Estevez N, Hebert J, Hwang J, Khaimskiy Y, Kim A, Lu Y, So PL, Tang X, Kohn MA, McCulloch CE, Kopelovich L, Bickers DR, and Epstein EH, Jr.** Basal cell carcinoma chemoprevention with nonsteroidal anti-inflammatory drugs in genetically predisposed PTCH1+/- humans and mice. *Cancer prevention research* 3: 25-34, 2010.

181. **Tatton-Brown K, and Rahman N**. Sotos syndrome. *Eur J Hum Genet* 15: 264-271, 2007.

182. **Teh MT, Gemenetzidis E, Chaplin T, Young BD, and Philpott MP**. Upregulation of FOXM1 induces genomic instability in human epidermal keratinocytes. *Molecular cancer* 9: 45, 2010.

183. **The Cancer Genome Atlas N, Genome sequencing centres: Washington University in St L, Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L, Mardis ER, Wilson RK, Genome characterization centres BCCA, Ally A, Balasundaram M, Butterfield YS, Carlsen R, Carter C, Chu A, Chuah E, Chun HJ, Coope RJ, Dhalla N, Guin R, Hirst C, Hirst M, Holt RA, Lee D, Li HI, Mayo M, Moore RA, Mungall AJ, Pleasance E, Gordon Robertson A, Schein JE, Shafiei A, Sipahimalani P, Slobodan JR, Stoll D, Tam A, Thiessen N, Varhol RJ, Wye N, Zeng T, Zhao Y, Birol I, Jones SJ, Marra MA, Broad I, Cherniack AD, Saksena G, Onofrio RC, Pho NH, Carter SL, Schumacher SE, Tabak B, Hernandez B, Gentry J, Nguyen H, Crenshaw A, Ardlie K, Beroukhim R, Winckler W, Getz G, Gabriel SB, Meyerson M, Brigham, Women's H, Harvard Medical S, Chin L, Park PJ, Kucherlapati R, University of North Carolina CH, Hoadley KA, Todd Auman J, Fan C, Turman YJ, Shi Y, Li L, Topal MD, He X, Chao HH, Prat A, Silva GO, Iglesia MD, Zhao W, Usary J, Berg JS, Adams M, Brooker J, Wu J, Gulabani A, Bodenheimer T, Hoyle AP, Simons JV, Soloway MG, Mose LE, Jefferys SR, Balu S, Parker JS, Neil Hayes D, Perou CM, University of Southern California/Johns H, Malik S, Mahurkar S, Shen H, Weisenberger DJ, Triche Jr T, Lai PH, Bootwalla MS, Maglinte DT, Berman BP, Van Den Berg DJ, Baylin SB, Laird PW, Genome data analysis: Baylor College of M, Creighton CJ, Donehower LA, Broad I, Getz G, Noble M, Voet D, Saksena G, Gehlenborg N, Dicara D, Zhang J, Zhang H, Wu CJ, Yingchun Liu S, Lawrence MS, Zou L, Sivachenko A, Lin P, Stojanov P, Jing R, Cho J, Sinha R, Park RW, Nazaire MD, Robinson J, Thorvaldsdottir H, Mesirov J, Park PJ, Chin L, Institute for Systems B, Reynolds S, Kreisberg RB, Bernard B, Bressler R, Erkkila T, Lin J, Thorsson V, Zhang W, Shmulevich I, Memorial Sloan-Kettering Cancer C, Ciriello G, Weinhold N, Schultz N, Gao J, Cerami E, Gross B, Jacobsen A, Sinha R, Arman Aksoy B, Antipin Y, Reva B, Shen R, Taylor BS, Ladanyi M, Sander C, Oregon H, Science U, Anur P, Spellman PT, The University of Texas MDACC, Lu Y, Liu W, Verhaak RR, Mills GB, Akbani R, Zhang N, Broom BM, Casasent TD, Wakefield C, Unruh AK, Baggerly K, Coombes K, Weinstein JN, University of California SCBI, Haussler D, Benz CC, Stuart JM, Benz SC, Zhu J, Szeto CC,**

Scott GK, Yau C, Paull EO, Carlin D, Wong C, Sokolov A, Thusberg J, Mooney S, Ng S, Goldstein TC, Ellrott K, Grifford M, Wilks C, Ma S, Craft B, Nci, Yan C, Hu Y, Meerzaman D, Biospecimen core resource: Nationwide Children's Hospital Biospecimen Core R, Gastier-Foster JM, Bowen J, Ramirez NC, Black AD, Xpath Error Unknown Variable Tname RE, White P, Zmuda EJ, Frick J, Lichtenberg TM, Brookens R, George MM, Gerken MA, Harper HA, Leraas KM, Wise LJ, Tabler TR, McAllister C, Barr T, Hart-Kothari M, Tissue source sites A-I, Tarvin K, Saller C, Sandusky G, Mitchell C, Christiana, Iacocca MV, Brown J, Rabeno B, Czerwinski C, Petrelli N, Cureline, Dolzhansky O, Abramov M, Voronina O, Potapova O, Duke University Medical C, Marks JR, The Greater Poland Cancer C, Suchorska WM, Murawa D, Kycler W, Ibbs M, Korski K, Spychala A, Murawa P, Brzezinski JJ, Perz H, Lazniak R, Teresiak M, Tatka H, Leporowska E, Bogusz-Czerniewicz M, Malicki J, Mackiewicz A, Wiznerowicz M, Ilsbio, Van Le X, Kohl B, Viet Tien N, Thorp R, Van Bang N, Sussman H, Duc Phu B, Hajek R, Phi Hung N, Viet The Phuong T, Quyet Thang H, Zaki Khan K, International Genomics C, Penny R, Mallery D, Curley E, Shelton C, Yena P, Mayo C, Ingle JN, Couch FJ, Lingle WL, Mskcc, King TA, Center MDAC, Gonzalez-Angulo AM, Mills GB, Dyer MD, Liu S, Meng X, Patangan M, University of California San F, Waldman F, Stoppler H, University of North C, Kimryn Rathmell W, Thorne L, Huang M, Boice L, Hill A, Roswell Park Cancer I, Morrison C, Gaudioso C, Bshara W, University of M, Daily K, Egea SC, Pegram MD, Gomez-Fernandez C, University of P, Dhir R, Bhargava R, Brufsky A, Walter Reed National Military Medical C, Shriver CD, Hooke JA, Leigh Campbell J, Mural RJ, Hu H, Somiari S, Larson C, Deyarmin B, Kvecher L, Kovatich AJ, Disease working g, Ellis MJ, King TA, Hu H, Couch FJ, Mural RJ, Stricker T, White K, Olopade O, Ingle JN, Luo C, Chen Y, Marks JR, Waldman F, Wiznerowicz M, Bose R, Chang LW, Beck AH, Maria Gonzalez-Angulo A, Data coordination c, Pihl T, Jensen M, Sfeir R, Kahn A, Chu A, Kothiyal P, Wang Z, Snyder E, Pontius J, Ayala B, Backus M, Walton J, Baboud J, Berton D, Nicholls M, Srinivasan D, Raman R, Girshik S, Kigonya P, Alonso S, Sanbhadti R, Barletta S, Pot D, Project team: National Cancer I, Sheth M, Demchok JA, Mills Shaw KR, Yang L, Eley G, Ferguson ML, Tarnuzzer RW, Zhang J, Dillon LA, Buetow K, Fielding P, National Human Genome Research I, Ozenberger BA, Guyer MS, Sofia HJ, and Palchik JD. Comprehensive molecular portraits of human breast tumours. *Nature* 2012.

184. **Tokes AM, Szasz AM, Juhasz E, Schaff Z, Harsanyi L, Molnar IA, Baranyai Z, Besznyak I, Jr., Zarand A, Salamon F, and Kulka J**. Expression of tight junction molecules in breast carcinomas analysed by array PCR and immunohistochemistry. *Pathol Oncol Res* 18: 593-606, 2012.

185. **Trempus CS, Morris RJ, Ehinger M, Elmore A, Bortner CD, Ito M, Cotsarelis G, Nijhof JG, Peckham J, Flagler N, Kissling G, Humble MM, King LC, Adams LD, Desai D, Amin S, and Tennant RW**. CD34 expression by hair follicle stem cells is required for skin tumor development in mice. *Cancer research* 67: 4173-4181, 2007.

186. **Tseng CP, Kim YJ, Kumar R, and Verma AK**. Involvement of protein kinase C in the transcriptional regulation of 12-O-tetradecanoylphorbol-13-acetate-inducible genes modulated by AP-1 or non-AP-1 transacting factors. *Carcinogenesis* 15: 707-711, 1994.

187. **Uren AG, Kool J, Berns A, and van Lohuizen M**. Retroviral insertional mutagenesis: past, present and future. *Oncogene* 24: 7656-7672, 2005.

188. **van Hogerlinden M, Rozell BL, Ahrlund-Richter L, and Toftgard R**. Squamous cell carcinomas and increased apoptosis in skin with inhibited Rel/nuclear factor-kappaB signaling. *Cancer research* 59: 3299-3303, 1999.
189. **Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, and Speleman F**. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology* 3: RESEARCH0034, 2002.
190. **Varley JM, Evans DG, and Birch JM**. Li-Fraumeni syndrome--a molecular and clinical review. *British journal of cancer* 76: 1-14, 1997.
191. **Venkatachalam S, Shi YP, Jones SN, Vogel H, Bradley A, Pinkel D, and Donehower LA**. Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation. *The EMBO journal* 17: 4657-4667, 1998.
192. **Wagner KU, McAllister K, Ward T, Davis B, Wiseman R, and Hennighausen L**. Spatial and temporal expression of the Cre gene under the control of the MMTV-LTR in different lines of transgenic mice. *Transgenic research* 10: 545-553, 2001.
193. **Wakabayashi Y, Mao JH, Brown K, Girardi M, and Balmain A**. Promotion of Hras-induced squamous carcinomas by a polymorphic variant of the Patched gene in FVB mice. *Nature* 445: 761-765, 2007.
194. **Walrath JC, Hawes JJ, Van Dyke T, and Reilly KM**. Genetically engineered mouse models in cancer research. *Advances in cancer research* 106: 113-164, 2010.
195. **Walter MJ, Payton JE, Ries RE, Shannon WD, Deshmukh H, Zhao Y, Baty J, Heath S, Westervelt P, Watson MA, Tomasson MH, Nagarajan R, O'Gara BP, Bloomfield CD, Mrozek K, Selzer RR, Richmond TA, Kitzman J, Geoghegan J, Eis PS, Maupin R, Fulton RS, McLellan M, Wilson RK, Mardis ER, Link DC, Graubert TA, DiPersio JF, and Ley TJ**. Acquired copy number alterations in adult acute myeloid leukemia genomes. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12950-12955, 2009.
196. **Wallace MD, Pfefferle AD, Shen L, McNairn AJ, Cerami EG, Fallon BL, Rinaldi VD, Southard TL, Perou CM, and Schimenti JC**. Comparative Oncogenomics Implicates the Neurofibromin 1 Gene (NF1) as a Breast Cancer Driver. *Genetics* 2012.
197. **Wang G, Williams G, Xia H, Hickey M, Shao J, Davidson BL, and McCray PB**. Apical barriers to airway epithelial cell gene transfer with amphotropic retroviral vectors. *Gene therapy* 9: 922-931, 2002.
198. **Wang GG, Cai L, Pasillas MP, and Kamps MP**. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nature cell biology* 9: 804-812, 2007.
199. **Wang GY, Wang J, Mancianti ML, and Epstein EH, Jr.** Basal cell carcinomas arise from hair follicle stem cells in Ptch1(+/-) mice. *Cancer cell* 19: 114-124, 2011.
200. **Wang NJ, Sanborn Z, Arnett KL, Bayston LJ, Liao W, Proby CM, Leigh IM, Collisson EA, Gordon PB, Jakkula L, Pennypacker S, Zou Y, Sharma M, North JP, Vemula SS, Mauro TM, Neuhaus IM, Leboit PE, Hur JS, Park K, Huh N, Kwok PY, Arron ST, Massion PP, Bale AE, Haussler D, Cleaver JE, Gray JW, Spellman PT, South AP, Aster JC, Blacklow SC, and Cho RJ**. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* 108: 17761-17766, 2011.

201. **Wang X, Levin AM, Smolinski SE, Vigneau FD, Levin NK, and Tainsky MA**. Breast cancer and other neoplasms in women with neurofibromatosis type 1: A retrospective review of cases in the Detroit metropolitan area. *American journal of medical genetics* 2012.
202. **Wang X, Mai TJ, Niu YN, Wang JW, and Guo YL**. [Identification of proteins interacting with androgen receptor- associated coregulator 267-alpha (ARA267-alpha) with the yeast two-hybrid system]. *Beijing da xue xue bao Yi xue ban = Journal of Peking University* 36: 514-518, 2004.
203. **Weber S, Niessen MT, Prox J, Lullmann-Rauch R, Schmitz A, Schwanbeck R, Blobel CP, Jorissen E, de Strooper B, Niessen CM, and Saftig P**. The disintegrin/metalloproteinase Adam10 is essential for epidermal integrity and Notch-mediated signaling. *Development (Cambridge, England)* 138: 495-505, 2011.
204. **Westfall MD, and Pietenpol JA**. p63: Molecular complexity in development and cancer. *Carcinogenesis* 25: 857-864, 2004.
205. **Wijnhoven SW, Zwart E, Speksnijder EN, Beems RB, Olive KP, Tuveson DA, Jonkers J, Schaap MM, van den Berg J, Jacks T, van Steeg H, and de Vries A**. Mice expressing a mammary gland-specific R270H mutation in the p53 tumor suppressor gene mimic human breast cancer development. *Cancer research* 65: 8166-8173, 2005.
206. **Woodworth CD, Michael E, Smith L, Vijayachandra K, Glick A, Hennings H, and Yuspa SH**. Strain-dependent differences in malignant conversion of mouse skin tumors is an inherent property of the epidermal keratinocyte. *Carcinogenesis* 25: 1771-1778, 2004.
207. **Xie J, Aszterbaum M, Zhang X, Bonifas JM, Zachary C, Epstein E, and McCormick F**. A role of PDGFRalpha in basal cell carcinoma proliferation. *Proceedings of the National Academy of Sciences of the United States of America* 98: 9255-9259, 2001.
208. **Xie J, Murone M, Luoh SM, Ryan A, Gu Q, Zhang C, Bonifas JM, Lam CW, Hynes M, Goddard A, Rosenthal A, Epstein EH, Jr., and de Sauvage FJ**. Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* 391: 90-92, 1998.
209. **Xu X, Wagner KU, Larson D, Weaver Z, Li C, Ried T, Hennighausen L, Wynshaw-Boris A, and Deng CX**. Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. *Nature genetics* 22: 37-43, 1999.
210. **Yamashita M, and Emerman M**. Retroviral infection of non-dividing cells: old and new perspectives. *Virology* 344: 88-93, 2006.
211. **Yang SH, Andl T, Grachtchouk V, Wang A, Liu J, Syu LJ, Ferris J, Wang TS, Glick AB, Millar SE, and Dlugosz AA**. Pathological responses to oncogenic Hedgehog signaling in skin are dependent on canonical Wnt/beta3-catenin signaling. *Nature genetics* 40: 1130-1135, 2008.
212. **Yant SR, Meuse L, Chiu W, Ivics Z, Izsvak Z, and Kay MA**. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nature genetics* 25: 35-41, 2000.
213. **Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, and Kay MA**. High-resolution genome-wide mapping of transposon integration in mammals. *Molecular and cellular biology* 25: 2085-2094, 2005.
214. **Yu Q, Geng Y, and Sicinski P**. Specific protection against breast cancers by cyclin D1 ablation. *Nature* 411: 1017-1021, 2001.

215. **Yuspa SH**. The pathogenesis of squamous cell cancer: lessons learned from studies of skin carcinogenesis. *Journal of dermatological science* 17: 1-7, 1998.
216. **Zen K, Yasui K, Gen Y, Dohi O, Wakabayashi N, Mitsufuji S, Itoh Y, Zen Y, Nakanuma Y, Taniwaki M, Okanoue T, and Yoshikawa T**. Defective expression of polarity protein PAR-3 gene (PARD3) in esophageal squamous cell carcinoma. *Oncogene* 28: 2910-2918, 2009.
217. **Zhang F, and Yu X**. WAC, a functional partner of RNF20/40, regulates histone H2B ubiquitination and gene transcription. *Molecular cell* 41: 384-397, 2011.
218. **Zhang Q, Sakamoto K, Liu C, Triplett AA, Lin WC, Rui H, and Wagner KU**. Cyclin D3 compensates for the loss of cyclin D1 during ErbB2-induced mammary tumor initiation and progression. *Cancer research* 71: 7513-7524, 2011.
219. **Zhao F, Chen Y, Zeng L, Li R, Zeng R, Wen L, Liu Y, and Zhang C**. Role of triptolide in cell proliferation, cell cycle arrest, apoptosis and histone methylation in multiple myeloma U266 cells. *European journal of pharmacology* 646: 1-11, 2010.
220. **Zhao Q, Caballero OL, Levy S, Stevenson BJ, Iseli C, de Souza SJ, Galante PA, Busam D, Leversha MA, Chadalavada K, Rogers YH, Venter JC, Simpson AJ, and Strausberg RL**. Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line. *Proceedings of the National Academy of Sciences of the United States of America* 106: 1886-1891, 2009.
221. **Zhao X, Malhotra GK, Band H, and Band V**. Derivation of myoepithelial progenitor cells from bipotent mammary stem/progenitor cells. *PloS one* 7: e35338, 2012.
222. **Ziegler A, Leffell DJ, Kunala S, Sharma HW, Gailani M, Simon JA, Halperin AJ, Baden HP, Shapiro PE, Bale AE, and et al.** Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers. *Proceedings of the National Academy of Sciences of the United States of America* 90: 4216-4220, 1993.