

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN  
ONCOLOGIA, EMATOLOGIA E PATOLOGIA**

Ciclo 35

**Settore Concorsuale:** 06/G1 - PEDIATRIA GENERALE, SPECIALISTICA E NEUROPSICHIATRIA  
INFANTILE

**Settore Scientifico Disciplinare:** MED/38 - PEDIATRIA GENERALE E SPECIALISTICA

**MODELING OF CLEAR CELL SARCOMA OF THE KIDNEY OF THE  
PEDIATRIC AGE**

**Presentata da:** Alberto Taddia

**Coordinatore Dottorato**

Prof.ssa Manuela Ferracin

**Supervisore**

Prof. Andrea Pession

**Co-supervisore**

Dott.ssa Annalisa Astolfi

**Esame finale anno 2023**



## Abstract

Clear cell sarcoma of the kidney (CCSK) is the second most common pediatric renal tumor, characterized in 90% of cases by the presence of internal tandem duplications (ITDs) localized in frame at the level of the last exon of *BCOR* gene. BCOR is a protein ubiquitously expressed in the human tissues, and constituting a core component of the non-canonical Polycomb Repressive Complex1 (PRC1.1), which performs gene silencing activity fundamental in pluripotency maintenance, differentiation induction and cell fate determination.

Internal Tandem duplications in the last BCOR exon at the level of PUF domain have been identified in many tumor subtypes and could affect PCGF1 binding and the subsequent PRC1.1 activity, although the exact oncogenic mechanism of ITD remains poorly understood due to the absence of a reliable model.

This project has the objective of investigating the molecular mechanisms underlying the oncogenesis of clear cell sarcomas of the kidney, approaching the study with different methodologies. A first cellular model in HEK-293 allowed to obtain important information about the functionality of BCOR, suggesting that the presence of internal tandem duplications generates an altered activity of the transcriptional repressor which is very different from a loss-of-function, as hypothesized so far in the literature. It has also been observed that the function of BCOR within the PRC1 complex varies with the variation of the ITDs present within the terminal part of the gene. Moreover, it allowed the identification of the molecular signatures evoked by the presence of BCOR-ITD, including its role in extracellular matrix interactions and invasiveness promotion.

The parallel analysis of WTS data obtained from 8 CCSK cases permitted the identification of a peculiar signature for metastatic CCSKs, highlighting a 20-fold overexpression of FGF3. This factor, indeed, promoted a significant increase in invasive ability in the in-vitro assays using the obtained cellular model in HEK-293.

In order to study the effects of BCOR-ITD over cell stemness maintenance and differentiation, an inducible model is being obtained in H1 cells. In this way, it will be possible to study of the functionality of BCOR-ITD in a context more similar to that of origin of the CCSKs, evaluating both the specific interactome and the phenotypic consequences caused by the mutation.



# Sommario

<b>1. INTRODUCTION</b> .....	1
<b>1.1 Pediatric Renal Tumors</b> .....	1
<b>1.1.1 Wilms' Tumor</b> .....	5
<b>1.1.2 Non-Wilms renal tumors</b> .....	9
<b>1.1.3 Clear cell sarcoma of the kidney</b> .....	10
<b>1.2 BCL6-Corepressor</b> .....	16
<b>1.2.1 Biological functions</b> .....	18
<b>1.2.2 Polycomb Repressive Complexes</b> .....	20
<b>1.3 Evolution of Genome editing technologies</b> .....	24
<b>2. AIM OF THE STUDY</b> .....	30
<b>3. METHODS</b> .....	31
<b>3.1 <i>In vitro</i> BCOR modeling</b> .....	31
<b>3.1.1 Generation of a stable BCOR-ITD model in HEK-293</b> .....	31
3.1.1.1 Single cell cloning .....	34
<b>3.1.2 PRC1.1 complex assembly evaluation</b> .....	34
<b>3.1.3 Model design in H1 cells</b> .....	35
<b>3.1.4 Flow cytometry analysis</b> .....	37
<b>3.1.5 Genomic DNA extraction</b> .....	37
<b>3.1.6 PCR screening</b> .....	38
<b>3.1.7 Gel extraction and Sanger sequencing</b> .....	40
<b>3.2 Fluorescence In Situ Hybridization</b> .....	42
<b>3.3 Transcriptome evaluation</b> .....	44
<b>3.4.1 Total RNA extraction</b> .....	44
<b>3.3.2 Real Time PCR</b> .....	44
<b>3.3.3 Whole Transcriptome Sequencing (WTS)</b> .....	46

<b>3.3.4 Bioinformatic Analysis</b> .....	47
<b>3.4 Cellular Assays</b> .....	47
<b>3.5 Protein identification</b> .....	48
<b>3.5.1 Co-Immuno Precipitation</b> .....	48
<b>3.5.2 Western blot analysis</b> .....	50
<b>4. RESULTS</b> .....	52
<b>4.1 Gene editing in HEK-293</b> .....	52
<b>4.1.1 Clones screening by flow cytometry analysis</b> .....	54
<b>4.1.2 Identification of editing events</b> .....	57
<b>4.1.3 Clones' characterization</b> .....	59
<b>4.1.4 Gene expression profiling</b> .....	63
<b>4.2 Characterization of metastasis-driving factors</b> .....	67
<b>4.2.1 Identification of a peculiar signature for metastatic CCSKs</b> .....	67
<b>4.2.2 In-vitro validation of the metastatic drive</b> .....	70
<b>4.3 PRC1.1 reconstruction</b> .....	71
<b>4.4 Inducible model design</b> .....	73
<b>5. DISCUSSIONS AND FUTURE PERSPECTIVES</b> .....	77
<b>6. BIBLIOGRAPHY</b> .....	84



# 1. INTRODUCTION

## 1.1 Pediatric Renal Tumors

The typologies of renal tumors that occur in the pediatric population are completely different from the ones that can be found in the adult age <sup>[1]</sup>. They constitute the IV group of the International Classification of Childhood Cancer, third edition (ICCC-3) <sup>[2]</sup> and represent 7% of the pediatric tumors in an age range between 0 and 4 years and the 11% between 15 and 18 years of age <sup>[3]</sup>.

Over a 20- years observation period (1975-95), the incidence was shown to be slightly higher in females and in black ethnicity <sup>[4]</sup>, while significantly lower in Asiatic countries <sup>[5]</sup>. This distribution suggests the relevance of predisposing genetic factors and is largely linked to Nephroblastoma's (or Wilms Tumor) epidemiological features. This histotype represent indeed around 85% of pediatric renal malignancies <sup>[6]</sup> however, thanks to significant developments regarding biopsy techniques, imaging, understanding of molecular biology and the drawing up of international protocols, different types of pediatric renal tumors have been defined (Table 1).

<b>Tumor type</b>	<b>Incidence</b>	<b>Age range</b>	<b>Peak of incidence</b>
<b>Wilms' Tumor</b>	85%	-	-
unilateral	80%	1 - 11 yo	3,5 yo
bilateral	4 - 7%	2 m - 2 yo	15 mo
<b>Clear cell sarcoma of the kidney</b>	3 - 4%	1 - 4 yo	2 yo
<b>Mesoblastic nephroma</b>	2 - 3%	0 - 1 yo	1 - 3 mo
<b>Rhabdoid tumor</b>	2 - 3%	6 m - 9 yo	6 - 12 mo
<b>Renal cell carcinoma</b>	2 - 3%	6 m - 60 yo	10 - 20 yo
<b>Miscellaneous</b>	~2%	-	-

*Table 1. Relative distribution by age and type of pediatric renal tumors, from Lowe L.H. et al. (2000).  
\* From Milar A.J. et al. (2017). yo= years old; mo=months*

In terms of incidence, the second most frequent pediatric renal malignancy is represented by the Clear cell sarcoma of the kidney (CCSK), followed by the Congenital mesoblastic nephroma (CMN), Rhabdoid tumor (RT), Renal cell carcinoma (RCC) <sup>[7]</sup> and other less frequent variants like those belonging to Ewings' sarcoma family.

Prognosis and clinical course of the different tumor types are extremely heterogeneous and depend mostly on the type and tumor stage at the time of diagnosis. The most appropriate diagnostic and therapeutic approach is defined in specific treatment protocols drawn up on previous results of national and international trials and studies. To date, the most relevant ones are the "UMBRELLA SIOP-RTSG 2016 protocol" conducted by the European "International Society of Pediatric Oncology" (SIOP) <sup>[8]</sup> and the "National Wilms' Tumor Study-5" (NWTS-5) in which, from the collaboration with the Children's Oncology Group, COG, the clinical protocols were organized according to risk classification in different trials (AREN-032B/0532/0533/0321/0534) <sup>[9]</sup>. Along with these groups are the "Associazione Italiana Ematologia e Oncologia Pediatrica" (with AIEOP TW 2003 protocols) and the "United Kingdom Children's Cancer and Leukemia Study Group" (UKCCSG, with UKW 3 protocols), which have contributed to expanding the amount of scientific acquisitions on the matter.

All the above-mentioned trials provide useful principles for a correct diagnostic classification of neoplasms, a staging system and a characterization of the necessary treatments, however they show discrepancies in terms of classification system for risk stratification the staging of the pathology (Table 2).

Stage	SIOP staging system	NWTS/COG staging system
<b>I</b>	<ul style="list-style-type: none"> <li>- The tumor is limited to the kidney or with fibrous pseudocapsule and is completely resected (resection margins clear)</li> <li>- The tumor may be protruding into the pelvic system (but not infiltrating their walls)</li> <li>- The vessels of the renal sinus are not involved</li> <li>- Intrarenal vessel involvement may be present</li> </ul>	<ul style="list-style-type: none"> <li>- Tumor is limited to kidney and completely resected</li> <li>- Renal capsule intact, not penetrated by tumor</li> <li>- No tumor invasion of veins or lymphatics of renal sinus</li> <li>- No nodal or hematogenous metastasis</li> <li>- No prior biopsy</li> <li>- Negative margins</li> </ul>
<b>II</b>	<ul style="list-style-type: none"> <li>- The tumor extends beyond kidney or the pseudocapsule, but is completely resected (resection margins clear)</li> <li>- The tumor infiltrates the renal sinus and/or invades vessels outside the renal parenchyma but is completely resected</li> <li>- The tumor infiltrates adjacent structures but is completely resected</li> </ul>	<ul style="list-style-type: none"> <li>- Tumor extend beyond kidney but is completely resected</li> <li>- Tumor penetrates renal capsule</li> <li>- Tumor in lymphatics or veins of renal sinus</li> <li>- Tumor in renal vein with margin not involved</li> <li>- No nodal or hematogenous metastasis</li> <li>- Negative margins</li> </ul>
<b>III</b>	<ul style="list-style-type: none"> <li>- Incomplete excision of the tumor</li> <li>- Any abdominal lymph nodes are involved</li> <li>- Tumor rupture before or intraoperatively</li> <li>- The tumor has penetrated through the peritoneal surface</li> <li>- Tumor thrombi presence</li> <li>- The tumor has been surgically biopsied prior to preoperative chemotherapy or surgery</li> <li>- Presence of necrotic tumor or chemotherapy-induced changes in a lymph node</li> </ul>	<ul style="list-style-type: none"> <li>- Residual tumor or nonhematogenous metastasis confined to abdomen</li> <li>- Any abdominal lymph nodes are involved</li> <li>- Peritoneal contamination or tumor implant</li> <li>- Tumor spillage of any degree occurring during or before surgery</li> <li>- Gross residual tumor in the abdomen</li> <li>- Biopsy of tumor (including fine-needle aspiration) prior to surgery</li> <li>- Resection margins involved by tumor or transection of tumor during resection</li> </ul>
<b>IV</b>	<ul style="list-style-type: none"> <li>- Hematogenous metastasis or lymph node metastasis spread beyond abdomen</li> </ul>	<ul style="list-style-type: none"> <li>- Hematogenous metastasis or spread beyond abdomen</li> </ul>
<b>V</b>	<ul style="list-style-type: none"> <li>- Bilateral renal tumors</li> </ul>	<ul style="list-style-type: none"> <li>- Bilateral renal tumors</li> </ul>

**Table 2.** Histopathological requirements and staging criteria, SIOP-NWTS/COG comparison. From Pater L. et al. (2020).

Treatment regimens are established based on whether the patient belongs to the categories shown in Table 2.

Outcomes for patients treated according to the COG or SIOP protocols overlap widely <sup>[10]</sup>, although survival outcomes vary significantly by risk classification. This is achieved through the evaluation of prognostic factors, determining the treatment according to the reference protocol and following the analysis of the tumor's histology and status.

The classification involves a schematization based on the morphological aspect, which is added to histological analysis, and more and more frequently an immunophenotyping or a genetic analysis of the mass. The latter allow a more effective recognition of cases on which interpretative doubts may arise, but also the possibility of recognizing diagnostic subgroups within a population.

The treatment generally involves an integrated approach between surgery, chemotherapy and radiotherapy, combined according to the specific protocol. Surgery consists of a radical nephrectomy and, where executable, remains the cornerstone of the therapy which allows to define the chemotherapeutic regimen and the eventual need for any radiotherapy <sup>[11]</sup>.

The chemotherapeutic approach change based on the selected protocol and consists of several cycles of poly-chemotherapy with variable combinations of vincristine, dactinomycin (actinomycin-D), doxorubicin (adriamycin), cyclophosphamide, etoposide, carboplatin and irinorecan based on the risk range membership <sup>[12]</sup>.

Radiotherapy treatment is not always carried out, but is recommended in case of advanced stages, the presence of post-surgical tumor residues and lymph node invasion, and generally involves post-operative irradiation.

The prognosis of childhood renal tumors is good, although long-term mortality due to toxicity from chemotherapy or second tumors still represent a serious problem. For this reason, all the protocols used aim at the progressive optimization of the treatments in order to reduce their toxicity. The frequency of tumor relapses not negligible too, and is higher than 25% in cases with an onset in an advanced stage and with unfavorable histology <sup>[13]</sup>.

### 1.1.1 Wilms' Tumor

#### *Epidemiology and clinical presentation*

Nephroblastoma, also known as Wilms' Tumor (WT), is the most frequent renal tumor and represents the fourth most common tumor during infancy. Under the age of 15 it affects one child over 10'000 with an average age at diagnosis of about 3.5 years <sup>[14]</sup> and a slight propensity for the female gender, and a progressive decline in incidence from 10 years of age. In adults (>18 yo) as well as infants, WT accounts for less than 1% of all the diagnosed kidney cancers <sup>[15]</sup>.

Although topographic variations have been noted, the incidence varies mainly according to different ethnic groups, with the highest occurrence frequency in Black Africans, while in Asians there is the lowest incidence.

Usually, it arises with a unilateral onset, while synchronous bilateral presentation affects less than 7% of cases <sup>[16]</sup>. Bilateral presentation is more frequently associated with the presence of predisposing hereditary mutations, often in the context of genetic syndromes, however familial Wilms tumors are rarely seen. Equally rare are the cases of extrarenal nephroblastoma, with an estimated rate of occurrence between 0.5 and 1% of all cases <sup>[17]</sup>.

The surgical approach plays an important role in the treatment of the disease, and in most of the cases requires the total removal of the affected kidney. However, the most promising results have been obtained with the innovation of the chemotherapeutic treatments, which generally consists of a double administration of vincristine and dactinomycin in case of localized tumors, and the addition of doxorubicin in presence of metastases.

The lungs are the most frequent site of metastasis, while bone and brain are unusual. Treatment of recurrences is often successful, provided that the tumor has a favorable histological profile and has not metastasized to the liver and mediastinum; however subsequent recurrences have a poor prognosis <sup>[18]</sup>.

Both adult and atypical nephroblastoma generally benefit from therapeutic approaches adopted for the classic forms.

### ***Pathological aspects***

During embryonic development, the fetal kidney develops from the ureteral bud, which forms the collecting ducts and from the metanephric mesenchyme or blastema. The collecting ducts develop in the renal pelvis and ureters, while the metanephric mesenchyme generates both the renal interstitium and, by means of the mesenchyme -epithelial transition process, also the epithelium of the glomeruli, proximal tubules and Henle Loops <sup>[19]</sup>.

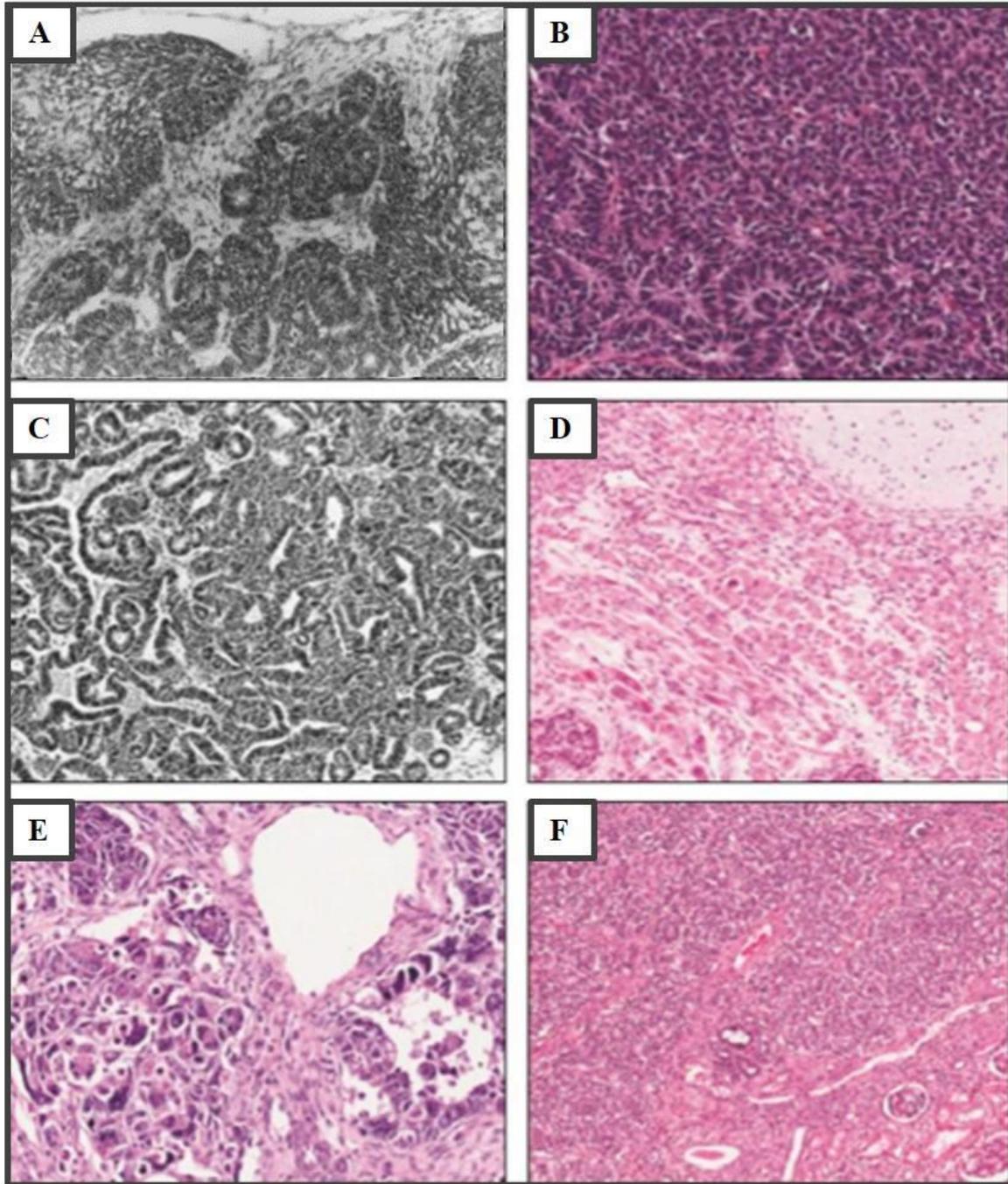
The blastema generally disappears by the 36<sup>th</sup> week of gestation, however, approximately 1% of newborns have a residual blastema at birth <sup>[20, 21]</sup>, the cells of which are termed as "nephrogenic remnants" (RN). RNs can undergo regression (even very late in adulthood), or neoplastic degeneration causing nephroblastoma, as are found in 30 - 44% of kidneys with WT <sup>[22]</sup>. "Nephroblastomatosis", a term introduced in 1961 by Hou and Holman <sup>[23]</sup>, defines the multifocal presence of RN and appears to be almost constant in cases of bilaterality <sup>[21]</sup>.

However, WT is not the only pediatric renal tumor associated with the presence of nephrogenic blastema, which can also be found in other predominantly benign neoplasms such as adenoma, adenofibroma and metanephric stromal tumor <sup>[24]</sup>.

Macroscopically, WTs appear as a solid, soft, gray and monomorphic asymptomatic mass, however tumors already treated with chemotherapy may appear hard, fibrotic, with necrotic-hemorrhagic areas <sup>[25]</sup>. Histologically, Wilms' Tumors have various forms. The classic presentation consists of a "triphasic" appearance which includes in variable proportions blastematous, epithelial and stromal components, up to cases of total absence of a component ("biphasic") or its exclusive presence ("monophasic") <sup>[22,6]</sup>. The presence of anaplastic areas as well as necrotic-haemorrhagic (especially if treated with chemotherapy) or cystic aspects are also possible.

The predominant appearance allows to predict the type of nephrogenic remains from which the neoplasm arose, and may reflect the genetic modifications that have occurred.

Blastematous-predominant tumors are the least differentiated and most aggressive, with a poor prognosis as they pose diagnostic challenges for being similar to other poorly differentiated tumors such as neuroblastomas, pNETs (primary neuroectodermal tumors) and synovial sarcoma. They are composed of layers of "small round blue cells" with a high nucleus/cytoplasm ratio, numerous mitotic figures and apoptotic bodies. These WT subtypes are characterized by an infiltrative growth pattern, resulting in 75% of cases in class III or IV staging <sup>[26]</sup>. However, the blastema proves to be the most responsive component to chemotherapy, leaving necrotic areas.



**Fig 1.** Wilms tumor, microscopic appearance. (A) Triphasic TW, taken from Charles A.K. et al. (1998); (B) Blastematosus TW, from Popov S.D. et al. (2016); (C) epithelial TW, courtesy of AFIP and PathologyOutlines.com; (D) Stromal TW, striated muscle, and cartilage differentiation taken from Popov S.D. et al. (2016); (E) Anaplastic TW, taken from Popov S.D. et al. (2016); (F) Nephrogenic remains in healthy cortical renal tissue, taken from Popov S.D. et al. (2016).

The dominant epithelial subtype can mimic various stages of tubular and glomerular development, as well as assume a heterologous differentiating epithelial morphology, such as squamous or mucinous epithelium. Paradoxically, this isotype is more chemoresistant and may require more intensive treatments than blastematosus tumors [22].

Lastly, Nephroblastomas with a stromal predominance can consist of both undifferentiated cells dispersed in a myxoid tissue, and areas of heterologous differentiation in muscular, adipose, bone or cartilage tissue [22]. This isotype must in some cases be included in the differential diagnosis with CCSK and mesoblastic nephroma, from which it can be distinguished by looking for foci containing other components [22].

The cystic appearance is a positive prognostic factor and falls within the risk stratification criteria used by the SIOP protocols for the attribution of the treatment regimen [27].

The presence of focal or diffuse anaplasia is a negative prognostic factor recognized by both the SIOP and COG protocols [27]. Approximately 5% of WT's have an anaplastic component, which occurs more commonly with increasing age of onset [28]. Anaplasia is not a histological condition but a cytological one, and all histological components can present anaplasia following the acquisition of more malignant molecular characters. The necessary criteria for the diagnosis of anaplasia are the presence of multipolar mitotic figures and large, hyperchromic nuclei [6].

Immunohistochemistry is useful in the analysis of preparations both in terms of differential diagnosis and in the recognition of histological subtypes, by means of the expression of specific tissue markers.

### ***Genetic aspects***

Several genes and molecular mechanisms are implicated in the pathogenesis of Nephroblastoma. Approximately 5% of WT patients have a predisposing genetic syndrome [29]. Among these, several depend on alterations of the *WT1* gene, coding for a transcription factor that plays a crucial role during embryogenesis. Microdeletions of *WT1* associated with alterations of the nearby *PAX6* gene determine the WAGR syndrome, with a more than 50% chance of generating WT. Nonsense or missense mutations in the *WT1* gene instead result in Denish-Drash syndrome, characterized by a very high risk of developing Wilms tumor (>90%). Mutations at the level of intron 9 may result in abnormal splicing resulting in Frasier syndrome [29].

Among the syndromes determining an increased risk of WT is also Beckwith-Wiedemann (BWS), caused by an altered genomic imprinting of the WT2 locus. This region is located on the p15.5 band of chromosome 11 and contains at least 10 genes normally expressed by only one of the two parental alleles. The apparently most involved gene is that of the growth factor *IGF2* (Insulin-like Growth Factor 2), causing an over-expression that determines tissue hypertrophy and tumor predisposition [29,30].

The genes implicated in these syndromes are also associated with the pathogenesis of sporadic WT [31]. In fact, an alteration of *WT1* is found in 10-20% of cases, while *IGF2* is deregulated in 69% of cases. Other genes implicated in pathogenesis include *MLLT1*, *MYCN* [32], as well as *WTX* and beta-catenin (*CTNNB1*) [33].

Numerous alterations have also been found in proteins implicated in RNA-mediated gene silencing processes, such as DROSHA, DICER and DIS3L2 [34].

In addition to the specific mutations, various chromosomal abnormalities consisting of aneuploidies, deletions and amplifications are also associated with the onset of WT. Among the most recurrent, chromosome 1q gain [35] and chromosome 1p and 16q loss [36] have been shown to carry a negative prognostic significance.

### 1.1.2 Non-Wilms renal tumors

The behavior of non-Wilms renal tumors is not sufficiently understood due to their heterogeneity and relative rarity; however, this group comprises a small portion of pediatric solid tumors and is associated with significant morbidity and mortality.

The tumors belonging to this group that appear with higher frequency are Clear cell sarcoma of the kidney (CCSK), Mesoblastic nephroma (MN), Rhabdoid tumor (MRT) and Renal cell carcinoma (RCC), as reported in Table 1.

### 1.1.3 Clear cell sarcoma of the kidney

Clear cell sarcoma of the kidney, indicated as CCSK (Clear Cell Sarcoma of the Kidney) is the second most frequent malignant renal tumor in childhood, representing about 5% of cases <sup>[37]</sup>, with 20 new diagnoses every year only in the USA.

Until 1970, CCSK was considered an aggressive variant of Wilms tumor <sup>[38]</sup>, from which it is still not distinguishable by clinical or radiological characterization <sup>[39]</sup>.

#### *Epidemiology and clinical presentation*

The neoplasm has an onset range from 2 months to 14 years of age, with an average age of 36 months, which appears to be the one with the best prognosis. Contrary to Wilms' tumor, which sees a slight propensity for the female gender, in CCSK there is a clear male predominance with a ratio of 2:1.

CCSK is managed with clinical protocols designed for Wilms' tumors and follows the same staging system.

Cases of CCSK on average are classified as follows: 27% of patients belong to stage I, 33% to stage II, 34% to stage III and 6% to stage IV. Only 4 cases with synchronous bilateral onset (stage V) <sup>[40,41]</sup> in patients with diffuse metastases have been reported in the literature.

The symptoms due to CCSK are completely overlapping to that of Wilms' tumor, mainly represented by abdominal distension and pain, the presence of a palpable mass and macroscopic hematuria.

The characteristic that led to the discrimination between the two tumor subtypes is the tendency of CCSK to metastasize to the bones. However, the main site of metastasis is represented by lymph nodes (59% of the cases), followed by bone (13%), lungs (10%) and liver (9%).

From an initial poor prognosis before the definition of CCSK as a separate entity and its distinction from Wilms' tumor, a marked increase in 6-year survival has been obtained by adding doxorubicin (Adriamycin) to the vincristine and dactinomycin usually used in chemotherapy treatments, despite the risk of cardiotoxicity, nephrotoxicity and related infertility problems <sup>[42]</sup>. Despite the results obtained, however, the relapse rate still affects 20% of cases <sup>[43]</sup>.

The new treatment regimens, consisting of the addition of cyclophosphamide (NWTS-5) and ifosfamide-etoposide-carboplatin (SIOP-2001) to the previous regimens, have further modified the timing and distribution of metastatic sites, bringing the mean time for the appearance of recurrences at 16-17 months, and reducing their frequency. Although the total number of metastatic cases is

decreasing, there is a statistically significant increase in secondary localizations to the brain, that apparently is not reached by current chemotherapy treatments.

### ***Pathological aspects***

Macroscopically, the tumor appears as a voluminous mass without a pseudocapsule but with fairly clear contours, which gradually replaces the healthy parenchyma. Diameter and weight are variable, but considerable sizes have been recorded (<25 cm, <1950 g). When cut, it is grey-beige in color, soft and with an abundance of mucoid material, mostly made up of hyaluronic acid.

Under microscopic observation, the classic pattern, present in 90% of cases, is characterized by an architecture made up of nests or cords of rounded cells, separated by regular fibrovascular septa with a chickenwire appearance <sup>[44]</sup>.

Tumor cells have round or oval and sometimes vesicular nuclei, with finely dispersed chromatin, inconspicuous nucleoli, and cytoplasm that cannot be stained with hematoxylin-eosin, which does not allow distinction between cell boundaries. The proliferative rate tends to be low, and in the vicinity of the capillaries the cells assume a fibroblast-like appearance, arranging themselves in cords. The extracellular matrix, rich in collagen and mucopolysaccharides, contributes to the "clear cell" appearance of the tumor.

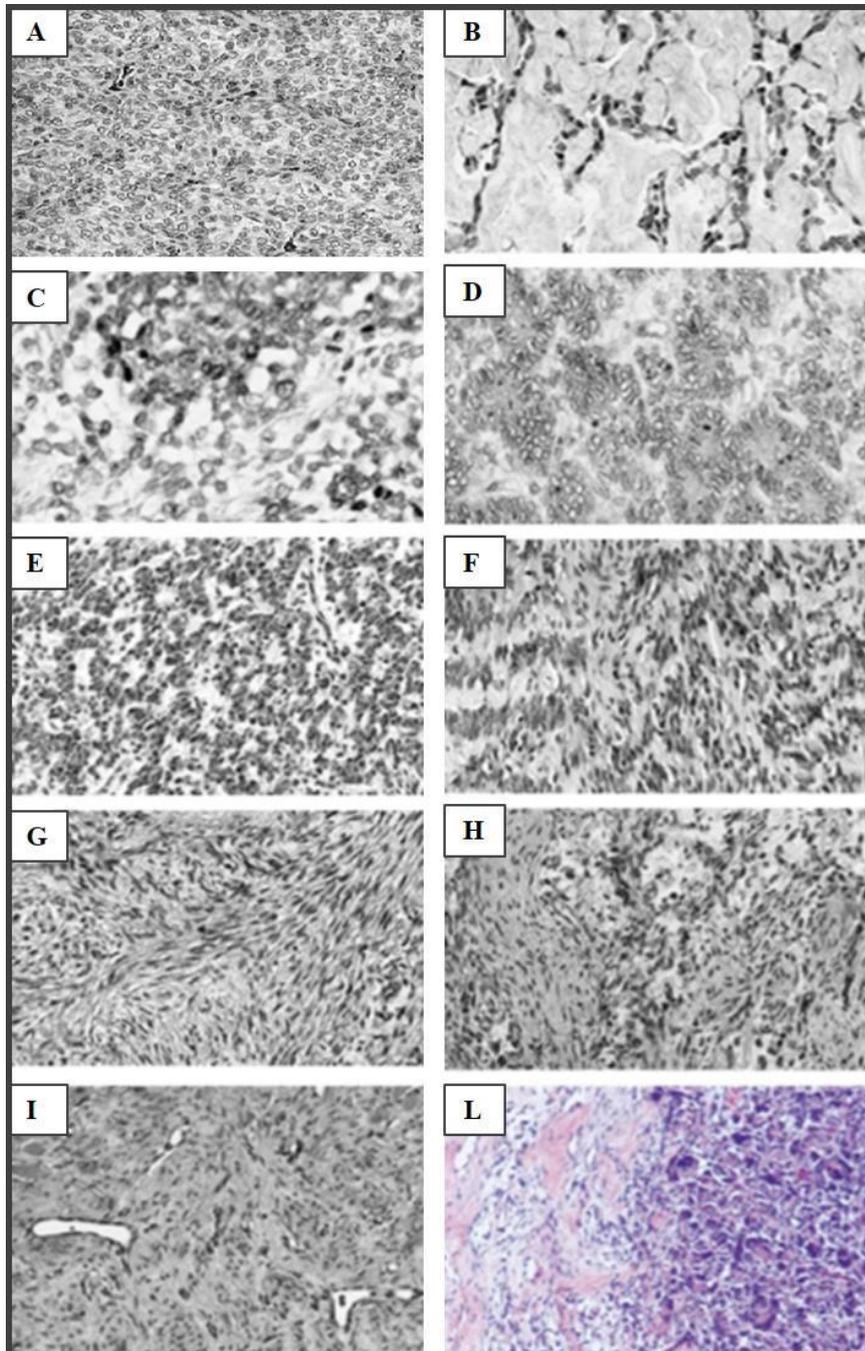
Although the classic pattern is relatively easy to diagnose, other patterns like epithelioid, spindle-shaped, sclerosing, myxoid, cystic, palisade, pericytomatous, and pleomorphic may complicate the diagnosis of the disease <sup>[45]</sup>.

None of the patterns listed above is highly represented in lymph node metastases, while recurrences more often assume sclerosing and hypocellular aspects, probably as a consequence of chemotherapy treatment <sup>[46]</sup>.

The histogenesis of CCSKs remains unknown, but appears to be unrelated to Wilms' tumors <sup>[47]</sup>. Immuno-histochemical analysis reveals a strong positivity for vimentin (marker of mesenchymal differentiation), often with a dot-like perinuclear distribution. The tumor also proves positive for the neuronal growth factor receptor (NGFR) and for Cyclin D1 <sup>[48]</sup>.

On the other hand, they are negative for desmin, EMA, CEA, cytokeratin, CAM 5.2, MIC-2, S100, NSE, c-KIT, CD34, synaptophysin, HMB-45 and WT1.

These features are extremely useful in distinguishing atypical and morphologically confusing patterns [25]. The small percentage of tumor cells positive for Ki-67 (MIB1 gene) staining suggest a low proliferative level.



**Fig 2.** CCSK, histological appearance of histological variants. (A) classic pattern; (B) sclerosing CCSK; (C) cellular CCSK; (D) epithelioid CCSK, acinar type; (E) Epithelioid CCSK, trabecular type; (F) palisading CCSK; (G) Storiform CCSK; (H) myxoid CCSK; (I) hypocellular CCSK; (L) CCSK with anaplastic focus (right). From Argani P. et al. (2000).

### *Genetic aspects*

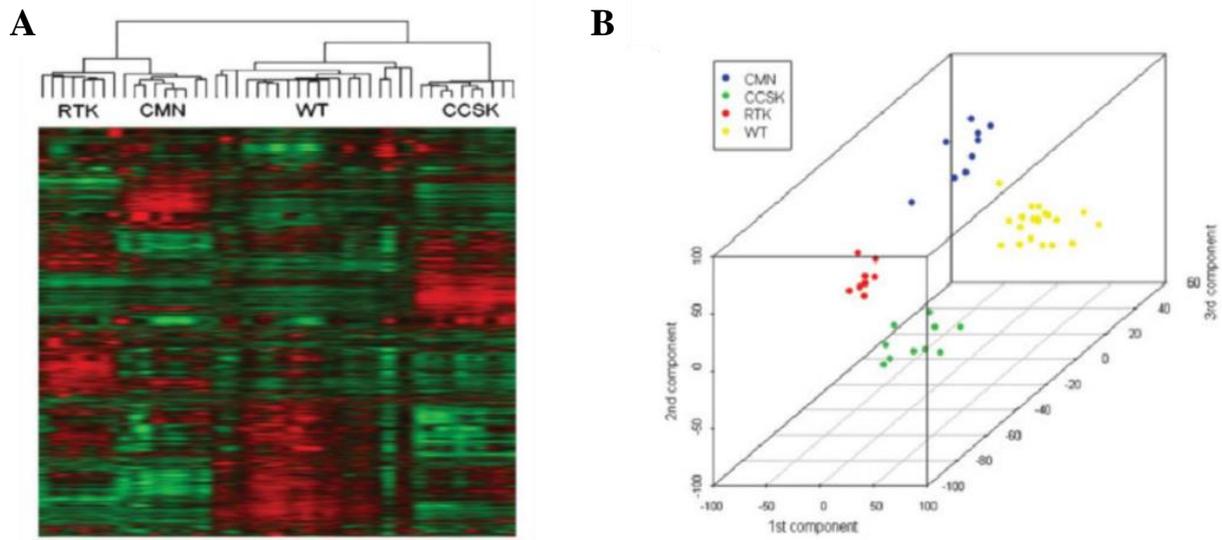
CCSK tends to be a genetically stable tumor that has a peculiar methylation profile <sup>[49]</sup>, and few genetic alterations have been found over the years.

The first mutation described is the YWHAE-NUTM2 fusion, found in a minority of cases <sup>[50,51]</sup>, while a small number of CCSKs have been shown to carry the BCOR-CCNB3 fusion gene, generated from the fusion of the transcriptional repressor *BCOR* (BCL-6 corepressor) and the *CCNB3* (Cyclin B3) genes <sup>[42]</sup>.

A decisive step forward in the study of this neoplasm has been made in recent years following the identification of a genetic aberration that presumably represents the oncogenic driver in the majority of CCSK <sup>[1,52,53]</sup>. Found in 80-100% of cases, this alteration is mutually exclusive with respect to the mutations listed above <sup>[54]</sup> and consists of different internal tandem duplications (ITDs) at the level of exon 15 of the gene coding for the transcriptional co-repressor *BCOR*.

These amplifications, found in both homozygous and heterozygous condition, have the characteristic of being in-frame and are associated with an overexpression of the BCOR protein, allowing important speculations on the origins of CCSK. Mutations affecting *BCOR* are not infrequent and have been characterized in a large series of neoplasms, in which, however, there is hypo-expression and/or loss of function.

In the attempt to identify a peculiar molecular signature for CCSK, unsupervised bioinformatic analysis of the CCSK transcriptional profile were conducted by several groups <sup>[55 - 61]</sup>, comparing the expression profiles from CCSK samples, healthy kidney tissue samples, as well as other kidney tumors. Those analysis showed the existence, among the various pediatric renal tumors, of clearly distinguishable and independent expression profiles (Fig. 3).



**Fig 3.** Pediatric renal tumor classification based on gene expression. (A) hierarchical clustering of normalized and log-transformed gene expression (obtained with RNA-expression-microarray with coverage of 11517 genes); in red the overexpressed genes, in green the under-expressed genes. (B) Principal component analysis (PCA) of gene expression in figure 3a. Abbreviations: CMN, cellular mesoblastic nephroma; WT, Wilms' tumor; CCSK, clear cell sarcoma of kidney; RTK, rhabdoid tumor of the kidney. Taken from Huang et al. (2006).

CCSK differentially expressed genes could largely be grouped into four categories: (a) a wide variety of neural markers, (b) members of the Sonic hedgehog pathway, (c) members of the phosphoinositide-3-kinase/Akt cell proliferation pathway, (d) pathways involved in embryonic development, including kidney formation.

- (a) Overexpression of genes involved in neural development (such as NPTX2, LHX2, ECEL1, IRX2, SATB2, NTRK3 and NGFR) involved in the oncogenesis of several neuroectodermal tumors, including medulloblastoma and pNETs;
- (b) Up-regulation of the Sonic Hedgehog pathway, fundamental for the embryonic morphogenetic processes of many organs and of laterality, as well as involved in the oncogenesis of some tumors, including basal carcinoma, medulloblastoma, rhabdomyosarcoma and pNETs. Among the overexpressed genes are found WNT11, GLI1, GLI2, WNT5A, WNT5B, PTCH1, SHH e FOXF1;

- (c) Up-regulation of the AKT/PI3K signaling pathway (to which are associated the already mentioned NTRK3 and NGFR, other than FOXO1A, FOXO3A, PDGFRA and PDGFA), involved in cell proliferation processes and in various neoplasms <sup>[57]</sup>;
- (d) Enrichment of pathways involved in embryonic development, including kidney formation. The principally identified markers as overexpressed in CCSK compared to fetal kidney are OSR1, FOXD1 and CITED1 <sup>[58]</sup>.

In addition, CD117 and epidermal growth factor receptor are up-regulated at the protein level in many CCSKs, providing potential therapeutic targets.

The presence of markers for two different developmental lineages in the embryonic kidney was investigated in order to identify the cellular origin of CCSK.

FOXD1, which identifies cells giving rise to stromal elements, and CITED1, a marker for cells primed for nephrogenic epithelial differentiation, were both highly expressed. In addition, the early embryonic marker OSR1 was expressed at higher levels in CCSK than in Wilms tumor, normal fetal kidney or adult kidney. As this marker discriminates the intermediate mesoderm from other mesodermal structures, it was therefore suggested that CCSK could arise from a mesodermal cell type that retains the capacity to initiate differentiation towards both nephrons and stroma, but remains locked in a primitive state.

EZH2 is overexpressed in many malignancies and its upregulation often correlates with increased aggressiveness <sup>[61]</sup>. Its overexpression has also been demonstrated in small cell lung cancer, another tumor which, like CCSK, is characterized by a strong hypermethylation of transcriptional promoters <sup>[62]</sup>. Furthermore, the data regarding the over-expression of EZH2 in CCSK can also be useful in the histological differential diagnosis with mesoblastic nephroma and above all with stromal-predominant nephroblastoma (par. 1.1.1), whose discrimination from CCSK can sometimes be complex <sup>[56]</sup>.

While BCOR over-expression is found only in ITD-BCOR positive CCSKs, little can be said about EZH2 up-regulation, which has indeed been reported to be present also in YWHAЕ-NUTM2B/E positive CCSKs (albeit in the total absence of quantitative descriptions). However, if this data were confirmed, it could be hypothesized that the up-regulation of EZH2 constitutes a pathogenetic scenario common to the two groups, although triggered through different cascades of events.

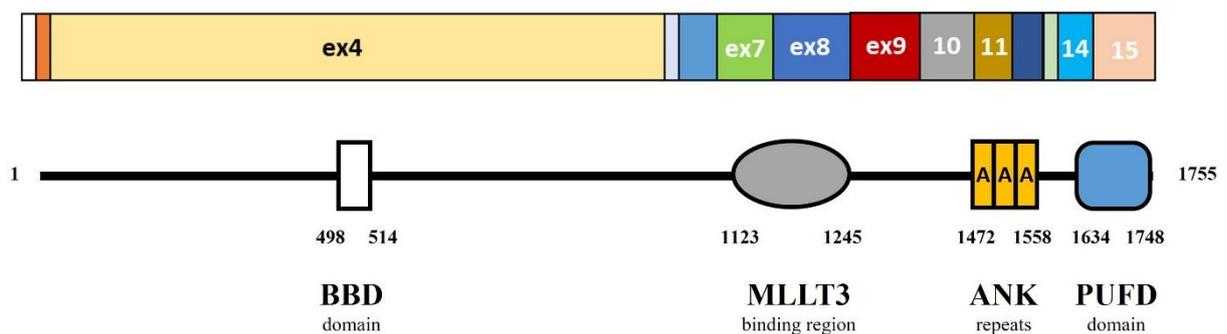
TCF21 hypo-expression in CCSK is probably attributable to the hypermethylation state of its promoter found in this tumor <sup>[59]</sup>. Its expression seems to be directly correlated to the lncRNA TARID expression, which is supposed as a crucial mediator in the demethylation of the TCF21 promoter <sup>[63]</sup>. In fact, consistently with this, TARID was found to be strongly down-regulated in the positive ITD-BCOR CCSKs. TARID is in turn regulated at the transcriptional level by the activity of PRC2 (of which it is a specific target) and its hypo-expression is therefore in line with the data concerning the over-expression of EZH2.

## 1.2 BCL6-Corepressor

BCL-6 corepressor, or *BCOR*, is a gene located on the short arm of the X chromosome (locus Xp11.4) <sup>[64]</sup> and has 15 exons coding for different protein isoforms (approximately between 1000 and 1750 aa), of which only some maintain the known interactions of the protein (Figure 3) based on the domains preserved by alternative splicing. The most known isoforms are the largest one, 1721 amino acids long, and the shortest isoform that only possesses the 1004 amino acids long amino-terminal region, and is known as BCOR-S (BCOR-Short).

Also on the X chromosome is the homologous gene BCORL1 (BcoR-like 1), with partially overlapping functional significance <sup>[65]</sup>.

### BCOR



**Fig 4.** Schematic structure and functional domains of BCOR. Adapted from Astolfi A. et al. (2019).

The BCOR protein is expressed in almost all tissues (unlike the BCOR-L1 paralogue, which is mainly expressed in the prostate and testis <sup>[65]</sup>), and can directly interact with:

- the repressive transcription factor BCL-6 (B-Cell Lymphoma 6) by means of the BBD domain (BCL-6 binding domain) on the amino-terminal side, replaced instead in BcoR-L1 by CBD, capable of binding the CTBP1 transcriptional co-repressor, in turn potentially able to bind Blc-6 <sup>[66]</sup>;
- PCGF1 and KDM2B proteins, by means of the carboxy-terminal PUF<sub>D</sub> (PCGF Ub-like fold discriminator) domain (last exon) and the nearby regions <sup>[67]</sup>, while direct interaction with other fundamental components of the PRC1.1 core, as RING1a/b, seems less probable <sup>[68]</sup>;
- the AF9 protein (MLLT3), by means of a domain conserved in some isoforms and located in the central portion of the protein <sup>[69]</sup>;
- the heat-shock protein HSPD1 interact with the N-terminal portion of BCOR to facilitate the protein folding and also the tri-methylation of histone 3. <sup>[70]</sup>
- probably the USP7 protein, again by means of the PUF<sub>D</sub> domain.

Through these interactions BCOR performs various repressive activities, both within protein complexes such as the non-canonical Polycomb Repressive Complex 1 (PRC1.1), and through its recruitment by BCL-6 via the BTB/POZ domain (Broad-complex, Tramtrack and Bric a brac / POxvirus and zinc finger). It is also possible that it acts as a steric inhibitor of AF9/MLLT3 (Trithorax transcriptional activator family), involved in the formation of the mRNA super-elongation complex (SEC) <sup>[69]</sup>.

BCL-6 is an important transcriptional repressor, mainly active in cells of the B-lymphocyte lineage (leading to the formation of the germinal center in the secondary follicle of peripheral lymphatic organs) and its dysregulation is a possible oncogenic factor in lymphomagenesis <sup>[71]</sup>. The action of this repressor is mediated by the interaction with several epigenetic regulatory factors such as N-CoR1, SMRT/N-Cor2 and precisely BCOR (via the POZ domain), as well as various HDACs and the NuRD complex <sup>[72]</sup>.

Through binding to its co-repressors, BCL-6 is able to recruit the repressive complex at the level of promoters of its target genes in lymphocytes (e.g. TP53, CCND1, PTEN, BMI1) <sup>[71]</sup>, but the functional importance of this mechanism of recruitment to other tissues is yet to be determined.

### 1.2.1 Biological functions

As can be understood from the complexity of the relationships undertaken, it is not yet possible to outline a precise picture of the biological functions of BCOR, however the fundamental role of this protein within the differentiation processes has emerged from cellular models.

Organisms that present BCOR knock-out or Loss-of-function mutations show alterations in the tissues generally involved in the oculo-facio-cardio-dental syndrome (OFCD), as well as noteworthy hematological anomalies, mainly consisting in lymphocyte depletion, confirming BCOR impact on BCL-6 activity <sup>[73]</sup>.

A loss of function at the level of the myeloid lineage instead seems to translate into increased proliferation <sup>[74]</sup>, while in stem cells it leads to growth arrest and differentiation <sup>[70]</sup>.

<b>Tumor family</b>	<b>Tumor hystotype</b>	<b>BCOR alteration</b>
Sarcomas	Clear cell sarcoma of the kidney	ITD-exon 15, BCOR-CCNB3 fusion
	Primitive myxoid mesenchymal tumor of infancy	ITD-exon 15
	Endometrial stromal sarcoma	ITD-exon 15, ZC3H7B-BCOR fusion
	Rhabdomyosarcoma	Frameshift insert/deletions, nonsense mutations
	Ossifying fibromyxoid tumor	ZC3H7B-BCOR fusion
	Undifferentiated sarcoma (small blue round or spindle cells)	BCOR-CCNB3 fusion, BCOR-MAML3 fusion, ITD-exon 15
CNS neoplasms	Medulloblastoma	Nonsense mutations, frameshift insert/deletions
	Retinoblastoma	Nonsense mutations, frameshift insert/deletions
	Gliomas, various histologies	Nonsense/missense mutations, frameshift insert/deletions
Hemolymphopoietic system tumors	Myeloid neoplasm	Nonsense/missense mutations, frameshift insert/deletions, splicing site
	Lymphoid neoplasm	Nonsense/missense mutations, frameshift insert/deletions, splicing site

Carcinomas	Salivary glands cancer	Missense mutations
	Adenoid cystic carcinoma	Missense mutations, frameshift insertions
	Endometrial carcinoma	Missense mutations, splicing site
	Gynecologic carcinosarcoma	Missense mutations
	Thymoma and thymic carcinoma	Missense mutations, frameshift insert/deletions

**Table 3.** Tumor isotypes in with identified BCOR mutations. Taken from Astolfi A. et al. (2019).

As far as the evaluation of the role of BCOR in tumorigenesis is concerned, the speculations on the matter are limited by the wide range of different associated neoplasms and somatic mutations found (Table 3).

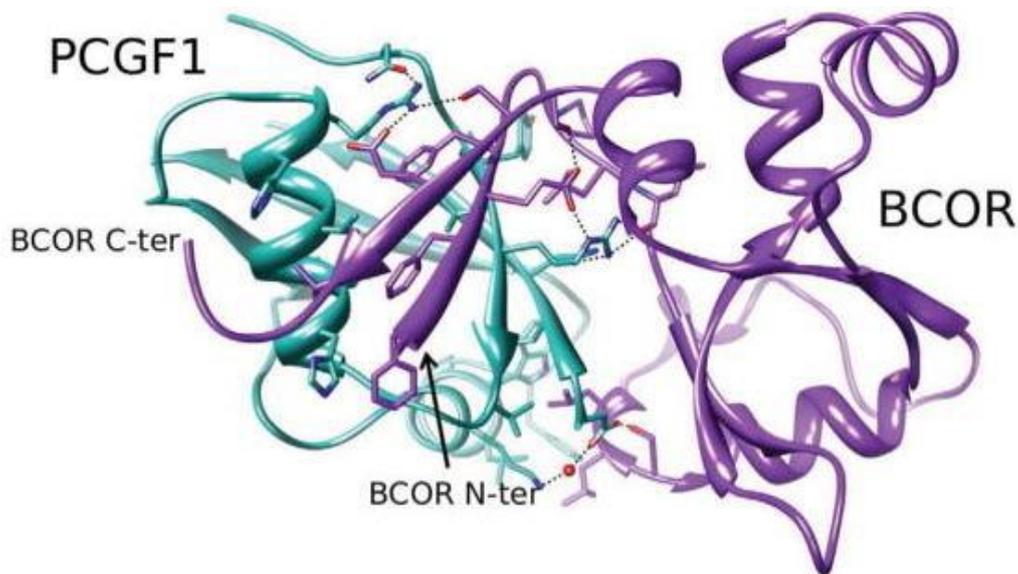
Indeed, in addition to CCSK, BCOR mutations have been described in pNET-related high-grade brain tumors (CNS-HGNET) [75], Endometrial stromal sarcoma (ESS) [76,77] and Endometrioid endometrial carcinoma (EEC) [78], some subgroups of undifferentiated / unclassified sarcomas (USS; Small Blue Round Cell and Spindle Cell variants) [79,80], Primary mesenchymal myxoid tumors of infancy [59,81], Rhabdomyosarcoma [82,83], Retinoblastoma [84], Medulloblastoma [85], as well as acute (and chronic) myeloid myelodysplasias and leukemias, especially with normal karyotype [86].

It should also be emphasized that, in the case of hematological neoplasms and in subgroups of Hepatocarcinoma, mutations of BCOR-L1 also often occur, albeit in lower percentages [87].

Starting from such a heterogeneous landscape, the comparison of the same mutations in different entities therefore becomes essential, and it could be assumed that also tumors characterized by the involvement of BCOR in the generation of fusion proteins may share some oncogenetic aspects with CCSK as they are often also associated with overexpression of the chimeric protein.

Numerous Internal Tandem Duplications (ITD) have been detected at the level of the last exon of the BCOR gene which seem to be related to the onset of CCSK (par. 1.1.3), and determine an elongation of the protein sequence between 22 and 60 amino-acids. The different ITDs, although variable, always include a minimum preserved residue of 13 amino acids. In several cases, short insertions of non-repeating nucleotide sequences were found at the end of the duplication, as well as, in a single case, the triplication of the sequence concerned.

All ITDs detected involve the PUF domain of BCOR, which is critical for binding to the RAWUL (RING finger and WD40-associated ubiquitin-like) domain of PCGF1 in the constitution of the PRC1.1 core (Figure 5).



**Fig 5.** *BCOR and PCGF1 interaction in the PRC1.1 core, 3D representation taken from Junco S.E. et al. (2013).*

### 1.2.2 Polycomb Repressive Complexes

The Polycomb Repressive Complexes (PRC), are multiprotein complexes consisting of different subunits involved in the repression of gene expression <sup>[88]</sup>.

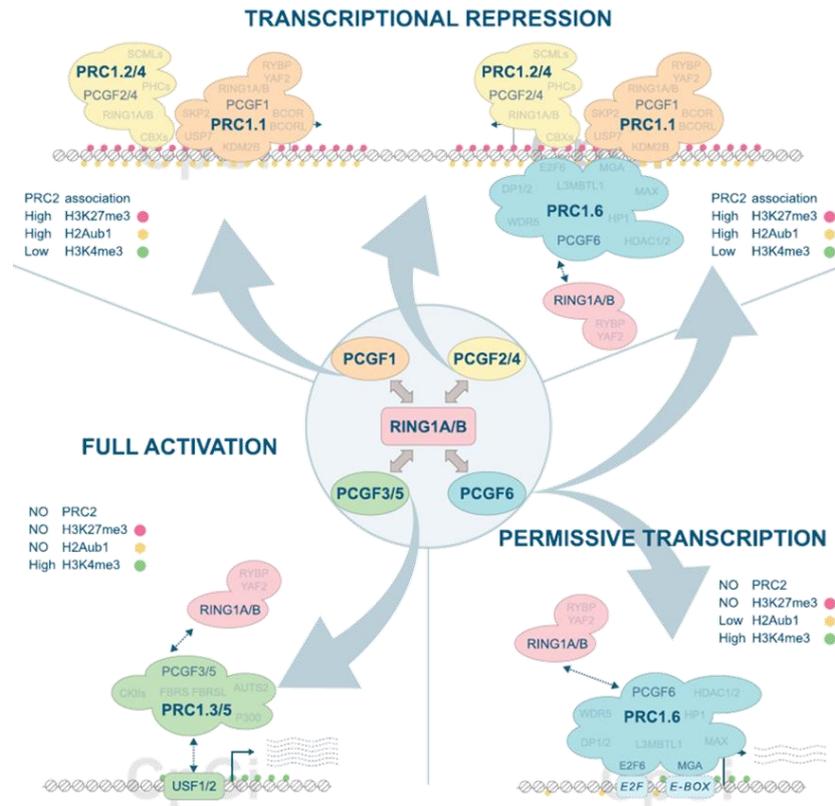
The first factors that were recognized as belonging to the Polycomb Group (PcG) were identified in *Drosophila melanogaster* where, together with the TrxG complex (Trithorax Group), they regulate the expression of the Hox genes, responsible for the correct anterior-posterior metameric development of the embryo <sup>[89]</sup>. The Polycomb complex owes its name to the discovery of a protein in male *Drosophila* mutants and refers to the presence of a greater number of sexual bristles than in the wild-type <sup>[90]</sup>.

PcG and TrxG, which respectively repress and activate the Hox genes during development, are recruited in *Drosophila* to specific DNA sequences referred to as PREs (Polycomb Responsive Elements). Although PcG and TrxG complexes are also conserved in mammals, PREs elements have not been identified in their genome. Despite this, recent studies have identified specific elements in human DNA capable of performing the same function as PREs, such as CpG islands. Binding to CpG islands allows for the regulation of many functions such as cell proliferation, stem cell identity, genomic imprinting, and X inactivation <sup>[91]</sup>.

The many functions of the Polycomb complexes can be broadly summarized into two specific epigenetic modifications, such as histones' methylation and ubiquitination <sup>[92]</sup>.

The cooperation between the two repressive complexes is essential in maintaining the pluripotency and plasticity of embryonic stem cells (ESCs), in which they localize at the promoters that regulate differentiation genes and prevent these cells from prematurely differentiating <sup>[93]</sup>.

Immunoprecipitation studies have shown that PcG proteins can be divided into two different complexes, called Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2), which act differently but in concert to maintain the repressive state and prevent transcription. Both PcG complexes also have ancillary subunits that determines the existence of many different PRC1 and PRC2 sub-complexes that possess specific molecular properties and biological functions <sup>[94]</sup>.



**Fig 6.** Schematic functions of the different PRC1 complexes. From Scelfo A. et al. (2019)

The number of Polycomb Group Proteins (PcG) genes found in mammals is more than double that of other species, so that more than 180 different PRC1s could be formed in humans if all combinations were tried. In reality, the complexes actually found in humans are currently a much smaller number and constitute the classic PRC1 and other so-called non-canonical PRC1s, named from PRC1.1 to 1.6 [95] (Figure 6).

The non-canonical complexes apparently play roles in the modulation of the expression of genes involved in metabolic or cell cycle-related processes, unlike classical PRC1, which has a prominent role in cell-type determination processes [96].

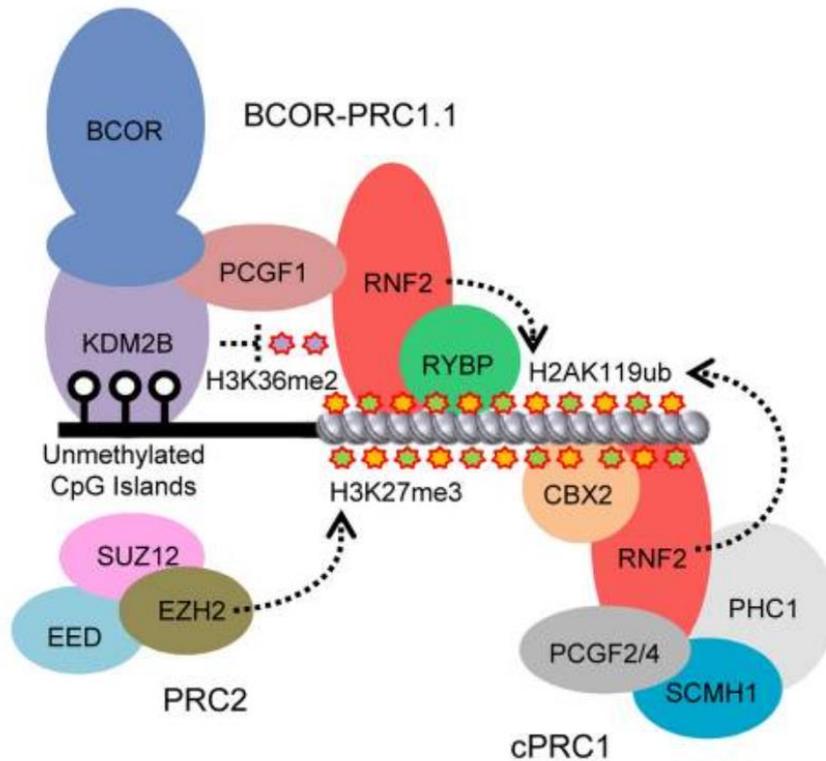
The PRC1 core is formed by the E3 ligases RING1A or RING1B that, by interacting with the products of one of the six Pcgf paralog genes (PCGF1-6), catalyze the mono-ubiquitination of histone H2A at lysine 119 (H2AK119ub1) [97,98,99]. The interaction with one of the PCGF paralogs is mutually exclusive and gives the name to the sub-complex.

PRC1.2 and PRC1.4 complexes are defined as canonical PRC1 (cPRC1) by the presence of CBX subunits that can bind H3K27me3, implying cPRC1 dependency on PRC2 activity. PRC1.1, PRC1.3,

PRC1.5, and the PRC1.6 forms exclude CBX proteins by associating with RYBP (or its paralog YAF2), do not recognize H3K27me3, and their activity is independent of PRC2. These PRC1 forms are defined as variant- or non-canonical PRC1 (vPRC1) and are tethered to target loci by intrinsic DNA binding activities. This includes PRC1.1 recognition of unmethylated CpG di-nucleotides by the KDM2B subunit <sup>[100]</sup>; PRC1.6 recognition of E-BOX and E2F DNA elements by the MAX/MGA and E2F6/DP dimers stably associated with the complex <sup>[101,102]</sup>; and PRC1.3 (and likely PRC1.5) by the recognition of an E-BOX variant directly bound by the USF1/2 transcription factors that can interact with and recruit the PRC1.3 complex to chromatin. Overall, this involves the cooperative activity of both cPRC1 and vPRC1 forms at repressed sites together with the exclusive presence of vPRC1 forms (PRC1.6 and PRC1.3) at several highly expressed genes. While PcG repressed loci display abundant H2AK119ub1 decoration, active vPRC1 targets are characterized by a low-to-absent H2AK119ub1 deposition <sup>[103]</sup>.

The PRC2 core is composed by two mutually exclusive methyltransferases, EZH1 and EZH2, that, by associating to the scaffold proteins SUZ12 and EED, catalyze mono-, di-, and tri-methylation of histone H3 lysine 27 (H3K27me1, H3K27me2, and H3K27me3) <sup>[104 - 107]</sup>. This complex exists in two major isoforms: PRC2.1 and PRC2.2.

PRC2.1 is characterized by the presence of polycomb-like subunits (PHF1, MTF2, and PHF19) that confers affinity of the complex to recognize unmethylated CpG islands, and either EPOP or PALI1 <sup>[108,109]</sup>. PRC2.2 is characterized by the AEBP2 and JARID2 subunits, where JARID2 provides affinity to PRC2.2 to bind directly to H2AK119ub1 <sup>[97, 110, 111]</sup>.



**Fig 7.** Representation of the proposed cooperative function of PRC complexes. Taken from Wang Z. et al. (2018).

Although the role of vPRC1 complexes in transcriptional regulation remains unclear, these observations suggest that H2AK119ub1 should play a major role in establishing transcriptional repression. Such a model implies an initial deposition of H2AK119ub1 that enhances PRC2 stability, H3K27me3 deposition, recruitment of cPRC1, and establishment of PcG repressive domains <sup>[112,113]</sup>.

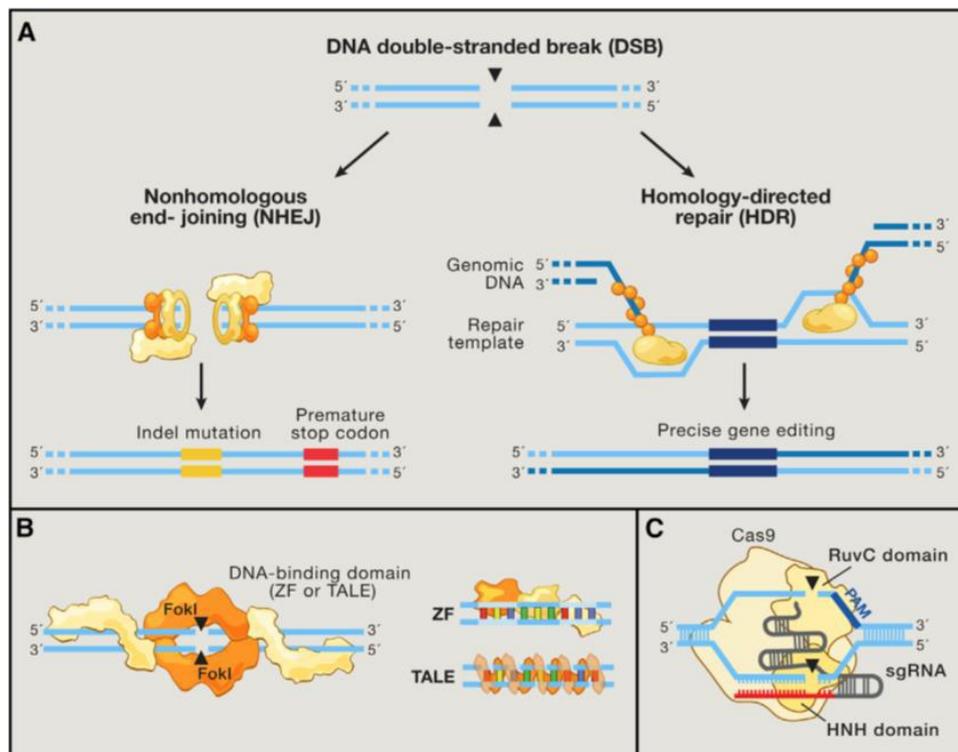
### 1.3 Evolution of Genome editing technologies

Genome-editing technologies based on programmable nucleases have significantly broadened our ability to make precise and direct changes in the genomic DNA of various species, including human cells. Indeed, with the means available, it is now possible to create a "knock-out" of entire genes, introduce mutations of one or more bases, or correct specific mutations in a gene therapy context <sup>[114,115]</sup>.

In the last 20 years, the discovery of 3 site-specific nucleases has led to a remarkable progress in genetic engineering. The first two nucleases are *Zinc Finger Nucleases (ZFNs)* and *Transcription Activator-Like Effector Nucleases (TALENs)*, although the most popular system nowadays in terms of simplicity and cost is the *Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9*.

The underlying principle of all three systems is a cut made by a nuclease in a precise point of the genome which can be followed by two types of repair <sup>[116]</sup>:

- Non-homologous-end-joining (NHEJ), leads to the introduction of aspecific sequences or deletions (Indel);
- Homologous recombination (HR), allows the integration into the genomic DNA of exogenous templates containing sequences of homology specific for a target site (Figure 6).



**Fig 8.** Genome editing technologies. (A) repair of DNA double strand breaks via nonhomologous end-joining (NHEJ) or homology-directed repair (HDR); (B) action of zinc finger proteins (ZF) and transcription-like effectors (TALE); (C) CRISPR/Cas9 action. Taken from Hsu P.D. et al. (2014).

Zinc finger nucleases (ZNFs) require specific peptides that can recognize certain nucleotide triplets of the target gene by homology (Figure 8b). The cut is performed by a bacterial endonuclease called FokI, which is activated only if in dimeric form and if the monomers of the enzyme are found on both DNA strands <sup>[117]</sup>.

The TALEN technology, similar to the one just described, still uses the FokI endonuclease, but unlike ZNF, the DNA sequences are not recognized as triplets but as single nucleotides thanks to the TALE proteins (Figure 8b). Transcription Activator-Like Effectors are transcription factors present in bacteria of the genus *Xanthomonas*, which bind DNA through domains formed by 33-35 amino acid repeats, which are highly conserved, with the exception of two variable nucleotides, modifiable according to the target sequence of the genome <sup>[118]</sup>.

Among the advantages of the TALEN technique, compared to ZNF, are a greater efficiency and specificity for the DNA sequence, with consequent reduction of off-target effects and toxicity.

Up to date, the CRISPR-Cas9 system is the fastest, simplest and cheapest method for genome editing. Its discovery, dating back to 1987, happened fortuitously during the studies of Nakatana and colleagues who noticed curious 29 nt repeats in the *E. coli* genome. Since then, a lot of information has been added regarding the functions of this unusual molecular mechanism.

Systematic analyzes of these sequences suggested extrachromosomal and phage-associated origin <sup>[119]</sup>, while further evidence led to speculation on immune memory function and defense mechanism, facilitated by high sequence variability.

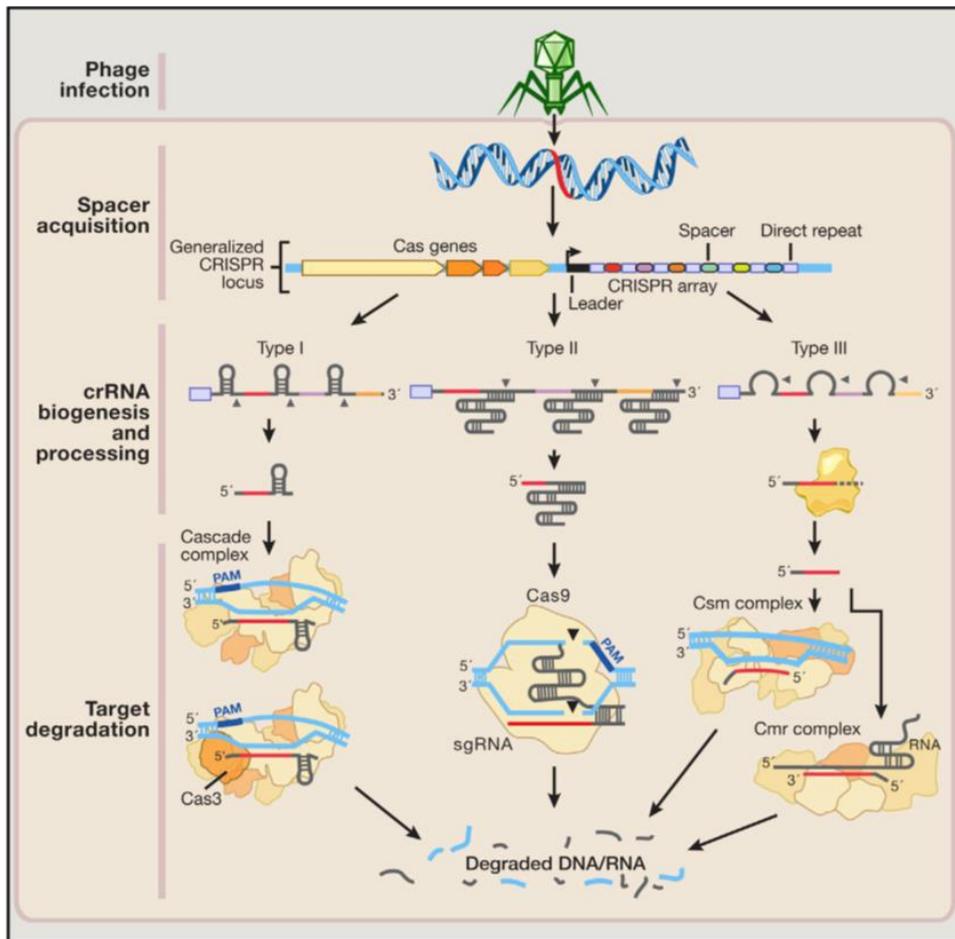
In 2002, the acronym CRISPR (clustered regularly interspaced short palindromic repeats) was coined to unify the descriptions of loci of bacterial genomes consisting of arrays of repeated sequences <sup>[120]</sup>.

CRISPR loci consist of a series of 29 nt direct repeats sandwiched by 32 nt variable sequences (spacers) corresponding to sequences contained in exogenous elements (protospacers).

Other sequences adjacent to the repeated elements of CRISPR sequences called CRISPR-associated (Cas) <sup>[120]</sup>, are well conserved and have allowed the identification of three types of CRISPR systems (types I-III) <sup>[121,122]</sup>.

While Cas genes are translated into proteins, most of the CRISPR-matching arrays (types I-III) are first transcribed as single-stranded RNA and then processed into shorter CRISPR RNAs (crRNAs) capable of directing the nucleolytic activity of certain Cas enzymes on target nucleic acids.

Type I and III CRISPR loci encode for several Cas proteins that will form complexes with various crRNAs to facilitate the recognition and destruction of nucleic acids [123,124]. In contrast, type II systems are associated with a relatively small number of Cas proteins (including Cas9).



**Fig 9.** CRISPR-associated (Cas) mechanisms of action in relation to different CRISPR types. From Hsu P.D. et al. (2014).

The first example of CRISPR-Cas9 function was based on the interaction between the Cas9 protein and a single small gRNA, a chimera consisting of the fusion of a crRNA homologous to the target region and a tracrRNA (trans-activating CRISP RNA) capable of folding in a hairpin structure required for interaction with Cas9 [125]. This interaction causes the cleavage of the double stranded DNA (DSB) in the region complementary to the crRNA portion. The cleavage is produced upstream of the protospacer adjacent motifs (PAMs) 5'-NGG/NCC-3', small sequences able to guide Cas9 in the cleavage [126,127]. Once the DSB is produced, the endogenous DNA repair systems (including HR and NHEJ) are activated using either a random insertion/deletion process (InDels) or other higher fidelity processes, respectively [128].

The first applications of the CRISPR-Cas9 system took advantage of the NHEJ as it is able to work simultaneously in different loci of the genome. In the absence of a template, this error-prone mechanism introduces random nucleotides that, if placed into the coding sequence of a gene, can cause that gene knockout or the alteration of the protein sequence.

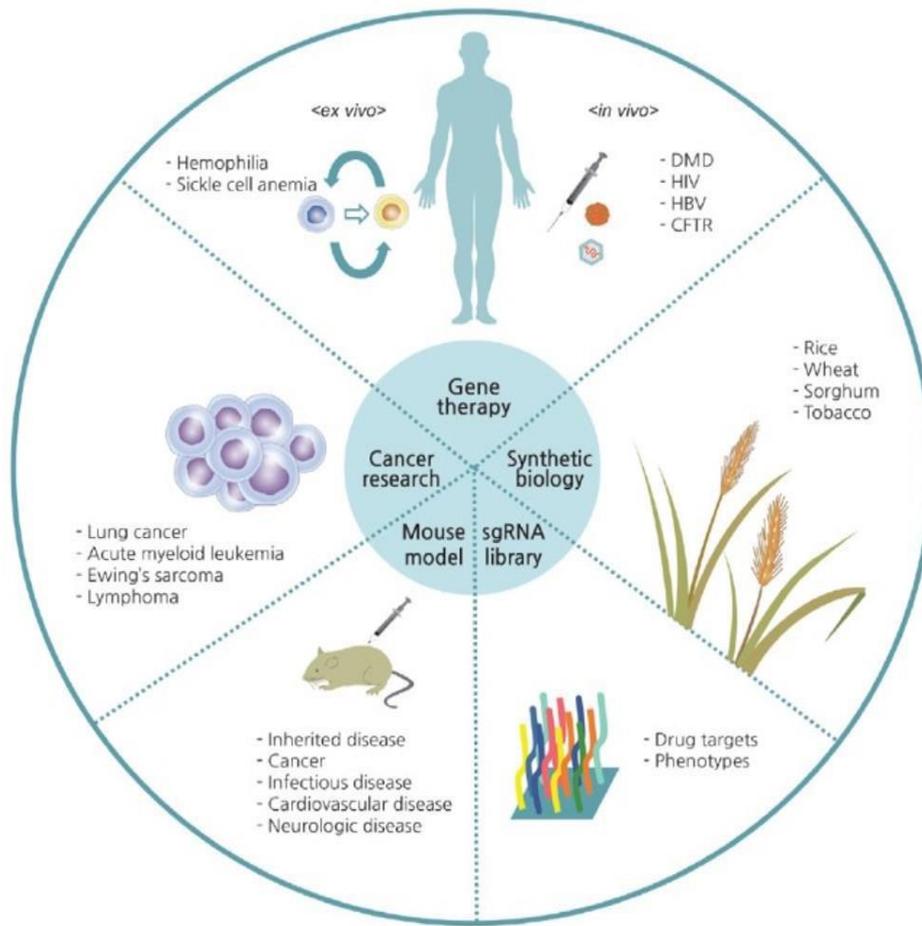
The homologous recombination (HR) repair system may be desirable in cases where accurate integration of exogenous genomic fragments such as reporter genes under the control of specific promoters is required.

In these cases, a DNA template (called donor vector) bearing the target sequence flanked by two homology arms is co-transfected together with the gRNA and the Cas9 coding sequence.

Given the great versatility, in recent years the applications of this system have extended to various model organisms, also increasing the need for high-throughput strategies to facilitate the selection of mutants quickly and cheaply.

The applications of the CRISPR/Cas9 system are many and range from oncological research to gene therapy for many monogenic diseases such as hemophilia, up to "drug discovery" and food biotechnology (Figure 9).

In the therapeutic field, new strategies are currently being studied for the correction of causal mutations (for example in Duchenne muscular dystrophy)<sup>[129]</sup>, the insertion of protective mutations (in the case of hemophilia and  $\beta$ -thalassemia) and the knock-out of genes encoding viral attack proteins (such as CCR5 for HIV attack). These genes can be corrected through an ex vivo approach in case the cell population is removable from the organism, otherwise strategies have been devised to perform in vivo gene editing<sup>[130]</sup>.



**Fig 10.** Possible areas of application of the CRISPR/Cas9 system in gene editing. Based on Kim E.J. et al. (2017).

The CRISPR/Cas9 system can also represent a powerful tool for the treatment of some neoplasms, by targeting the main oncogenes <sup>[131]</sup>, but at the same time it represents a specific mechanism for the introduction of mutations into target genes to generate isogenic preclinical models of disease both in the oncological and in other fields of human pathologies.

## 2. AIM OF THE STUDY

Clear Cell Sarcoma of the Kidney is a rare pediatric neoplasia characterized by the presence of different internal tandem duplications (ITD) in the last exon of the *BCOR* gene.

These duplications are strictly in-frame and associated with an overexpression of the mutated protein which allowed important speculations on this tumor's origins, however there are no documented in-depth studies either on the impact of this alteration on BCOR functions nor on the repressor complexes with which it interacts.

The primary objective of this research project is to characterize in-depth the oncogenic role of the Internal Tandem Duplication of BCOR through the development of a cellular model that reliably reproduces the biological and molecular features of Clear Cell Sarcoma of the Kidney. Two different but complementary approaches were designed: a model based on the HEK-293 cell line and a second one that involves the use of the embryonal stem cell line, H1.

The HEK-293 cell line represents an optimal model due to its embryonic renal derivation, and served as a pre-emptive model to investigate the molecular mechanisms underlying the tumorigenic potential of BCOR-ITD.

In order to study the effects of this genetic anomaly over cell stemness maintenance and differentiation, an inducible model is being obtained in H1 cells, in order to study both its growth in 2-D cultures and during the development of renal organoids.

In parallel, whole transcriptome sequencing data of metastatic and localized BCOR-ITD positive CCSK at diagnosis were compared in order to find specific differential signatures that could implement the identification of more aggressive tumour entities.

## 3. METHODS

### 3.1 *In vitro* BCOR modeling

Three different approaches were used in order to analyze BCOR-ITD functions.

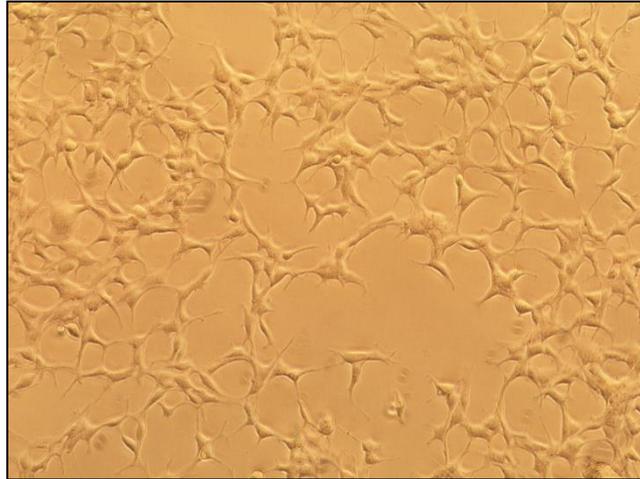
A first *in-vitro* model in HEK-293 was used to carry on preliminary investigation on this mutation's function in a cellular contest, while a second inducible model in embryonal stem cells is still being generated. Along with the recombinant cell models, a transient overexpression in HEK-293 FT of the PRC1.1 core components was used to assess the binding capacity of BCOR-ITD.

#### 3.1.1 Generation of a stable BCOR-ITD model in HEK-293

The cell line chosen as a starting point for the development of an *in vitro* model to study the CCSK was the HEK-293 line. Its derivation from human embryonic kidney cells makes this cell line a good candidate as a preliminary cellular model to study this pathology. HEK-293 cells present an hypotriploid karyotype with, notably, three copies of the X chromosome.

Preliminary tests confirmed BCOR expression both at a transcriptional and protein level along with the absence of mutations, especially in the last exon.

The cells were grown in DMEM with 10% FBS (*Fetal Bovine Serum*, Euroclone), 20  $\mu$ M L-Glutamine and 1% Penicillin - Streptomycin mix (Gibco) and splitted every 4 days with trypsin 1X (Gibco).



**Fig 11.** HEK-293 cells. Microphotography with 10X magnification.

The donor vector was obtained by linking specific BCOR regions from CCSK tumor sample (previously characterized in Astolfi et al.2015) with P2A-GFP sequences from previously obtained vectors via fusion PCR. The construct was then cloned into a pGL3 basic vector (Addgene).

The primers used to amplify the specific fragment are reported in Table 4.

<b>Primer</b>	<b>Sequence</b>
BCORint14/15 F2	GCCTCACCTTAGAAAGGAATGG
BCOR EX 15 REV CL	AAGTTTGTGCGCCGGATCCCCAGTAATTG
P2A-GFP F2	CTGGCCTCAGACAATTACTGGGGATCCGGC
P2A-GFP R2	CATGGTGGGTCCACCTTGCTTATCTGTACA
GFP F1	CTGCTAACCATGTTTCATGCC
GFP R3	GAGGGAAAAAGATCTCAGTG
BCOR UTR F2	CATGGACGAGCTGTACAAGTAAGCAAGGTG
BCOR UTR R2	CATTCCAAGCAACCTCAACATAT
BCOR UTR+PURO2F	GCAAGCCCGGTGCCTGAGCAAGGTGGACCC

**Table 4.** Primers used for fusion PCR

The day prior to the transfection,  $5 \times 10^5$  cells in 2 mL of media were plated into a 6 well plate (Corning) in order to reach around 70% confluence.

The day after, the media was replaced with 1,75 mL of fresh media and the cells were treated with 1  $\mu\text{g}$ /well of Donor and guide vectors in the following combinations:

- Donor vector + gRNA1 vector (G1)
- Donor vector + gRNA2 vector (G2)
- Donor vector + Scramble control vector (SCR)

The Donor vector obtained by inserting BCOR exon 15 carrying a 90 bp ITD, followed by the P2A linker upstream of GFP and BCOR 3'UTR, in a pGL3basic plasmid. The gRNA vectors G1 and G2 were carrying two different spacer sequence specific for BCOR last exon for the Cas9 binding in a PX459 plasmid.

Every combination of plasmids was diluted in 250  $\mu\text{L}$  of media. The transfection was carried out with the lipofection technique, using 12  $\mu\text{L}$  of Lipofectamine 2000 reagent (Life Technologies) and following the manufacturer's protocol.

For each lipofection, 12  $\mu\text{L}$  of reagent were diluted in 250  $\mu\text{L}$  of media and were added to 150  $\mu\text{L}$  of the plasmids' dilution prior to a 20 min incubation at room temperature. Each reaction was then added to their respective wells dropwise. One well received only the Lipofectamine 2000 dilution, while the remaining two wells were used as negative controls (NT). The plate was put in the cell culture incubator at 37°C. The target sequences of the two guides are shown in the table below.

Guide	Sequence
BCOR G1	ATCTGGCCTCAGACAACTAC
BCOR G2	GGGGTGGAGCCACTCTACAG
SCR	GCACTACCAGAGCTAACTCA

**Table 5.** *BCOR-specific and scramble spacer sequences inserted in the guide vectors.*

After 24 hours, transfected cells were selected for the following 48 hours by adding Puromycin in the media at the final concentration of 3 µg/mL. The surviving cells were expanded in normal media and then the two groups transfected with the Donor vector were sorted for GFP expression.

#### *3.1.1.1 Single cell cloning*

Both the sorted cells and the SCR cells were expanded in 25 cm<sup>2</sup> flasks and when splitted, a portion of the cells was diluted via serial dilutions to the final concentration of 0,5 cells/100 µL in 10 mL of fresh media. For each experimental group four different 96 well plates have been plated with 100 µL of cellular suspension for each well.

After 48 h every plate was evaluated to check for single cells in the wells. Those with multiple cells were marked so as not to be taken into consideration later. The plates were left in the incubator for the next 7 days and then every clone was expanded further.

Every clone obtained this way from transfection with the guides vector was then evaluated again for GFP expression by flow cytometry, and only a few clones were selected for further characterizations.

#### **3.1.2 PRC1.1 complex assembly evaluation**

The cell line used were the Hek-293 FT, cells derived from human embryonal kidney cells transformed with the SV40 large T antigen in order to be used for viral packaging.

The cells were cultured in DMEM with 10% FBS, 1% of penicillin-streptomycin mix, Non-Essential Aminoacids, Sodium Pyruvate, GlutaMAX (Thermo Fisher), and 5 µM β-Mercaptoethanol. The cells were grown in 15 cm dishes and splitted every 4 days with tryPLE 1X (Gibco). The day prior to transfection, 21\*10<sup>6</sup> cells were plated in gelatin-coated plates and incubated overnight.

The plasmids for transient overexpression of the PRC1.1 complex core components were mixed with one of the BCOR variants in a fresh tube as reported in Table 6 and incubated at 65°C for 15 min.

Plasmid	Size (Kb)	µg
BCOR WT	220	6
BCOR-ITD1	220	6
BCOR-ITD2	220	6
BCOR-ITD3	220	6
BCOR ΔPUFD	220	6
KDM2B	150	4.05
RING1B	40	1.08
PCGF1	25	0.605
RYBP	30	0.81
SKP1	20	0.54

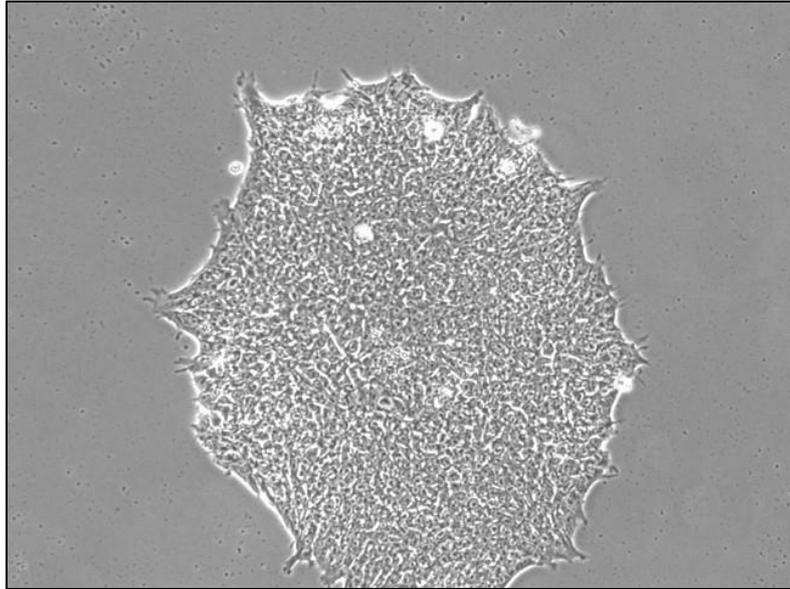
**Table 6.** *Plasmids for transient expression of the PRC1.1 core factors*

Each plasmids mixture was diluted in 2 mL of fresh DMEM, then 70 µL of PEI (Thermo Fisher) were added to each tube, mixed by briefly vortexing and incubated for 10 min at room temperature.

After the incubation, the mixture was added dropwise to the cells and the plates were then left in the incubator for 96 hours before being used for the co-immunoprecipitation protocol.

### 3.1.3 Model design in H1 cells

H1 cells were cultured in StemFlex media (Gibco) on Matrigel-covered (Gibco) 6 well plates. For standard passaging, cells were incubated for 1 h with Dispase (Sigma-Aldrich), washed with DMEM F-12 (Gibco) and splitted in a new matrigel-coated plate. For procedures requiring the single cell condition, cells were detached in 500 µL of Accutase (Thermo Fisher) after a 5 min incubation. Cells were then centrifuged for 5 min at 1100 rpm and resuspended in StemFlex supplemented with ROCK inhibitor Y27632 10 µM.



**Figure 12.** *H1 cells. Microphotography with 10X magnification*

The multi-step design for the BCOR-ITD-inducible model required and will require multiple steps of recombination through both TALENs or CRISPR/Cas9 system, using the *Neon Transfection system* (Thermo Fisher).

The insertion of a tamoxifen-inducible CRE-ERT2 chimeric gene in the AAVS1 locus was executed by co-transfection of AAVS1-TALEN-L and AAVS1-TALEN-R plasmids (Addgene) with a donor vector carrying the chimeric gene and the puromycin resistance cassette downstream of the CMV enhancer.

To insert the 3XFlag – HA Tag in the N-terminal region of BCOR, the cells harbouring the selected Cre-ERT2 construct were co-transfected with a guide vector carrying the Cas9 sequence under a constitutive promoter and a donor plasmid reporting a sequence specific for the second exon of BCOR with a donor plasmid carrying the Neomycin resistance cassette and the HSV TK “suicide” cassette along with the 3XFlag-HA tag, located between the 5’ UTR and the start of the second exon of BCOR.

For each nucleofection the needed plasmids were mixed together into a fresh tube and incubated at 65°C for 15 min beforehand, then kept on ice until use. Meanwhile,  $1 \times 10^6$  cells for each reaction were detached in single cell and pelleted after a wash in PBS. The cells were resuspended with 110µL of plasmids resuspended in ice-cold buffer R, electroporated with one pulse at 1050 V for 30 ms and then plated in a freshly coated 6 well plate.

The cells were let grow for 48 h and then selected for the specific antibiotic resistance for 3 to 6 days (50 µg/mL Neomycin or 0,5 µg/mL Puromycin), then the single colonies were picked and expanded before being tested for the correct insertion.

### **3.1.4 Flow cytometry analysis**

Flow cytometry is a technique that allows to analyze and quantify physical and fluorescent parameters and distinguish different cellular populations inside the same sample.

To analyze the GFP expression by flow cytometry (BD FACSCanto II), around  $10^6$  cells from each clone were detached and collected into 5 mL tubes. Every sample was washed twice with PBS 1X (Phosphate Buffered Saline) to eliminate traces of the media that could interfere with the readings.

For intracellular staining cells were harvested, washed in PBS then fixed and permeabilized in ice-cold 90% methanol overnight. Then fixed cells were centrifuged, washed in PBS supplemented with 4% FBS and incubated with primary rabbit anti-BCOR antibody (ab135801, Abcam). After one-hour cells were centrifuged and incubated with FITC-conjugated goat-anti-rabbit secondary antibody and analyzed through flow cytometer.

The data obtained from flow cytometry assays were analyzed with the FlowJo software (TreeStar).

### **3.1.5 Genomic DNA extraction**

Genomic DNA was extracted with *QIAmp Mini kit* (Qiagen). Around  $10^6$  cells were detached and centrifuged at 1200 rpm for 5 min. The cellular pellet was resuspended in 200 µL of PBS 1X and incubated with 200 µL of AL buffer and 20 µL Proteinase K at 56°C for 10 min to lysate the cellular components. At the end of the incubation, 200 µL of Ethanol (96-100%) were added to each sample and mixed vigorously before transferring them to the columns (*QIAmp spin column*). The columns with the lysates were centrifuged at 8000 rpm for 1 min to allow DNA adsorption on the filter columns, while the other cellular components were eliminated. To eliminate the majority of the cellular residues the columns were washed once with 500 µL of AW1 buffer and

then with 500  $\mu\text{L}$  of AW2 buffer, centrifuging every time at 8000 rpm for 1 min. At last, the samples were centrifuged for 1 min at 14000 rpm to get rid of ethanol residues.

The obtained genomic DNA was eluted in 100  $\mu\text{L}$  of AE buffer and centrifuged at 8000 rpm for 1 min. The flow-through was evaluated qualitatively and quantitatively using a NanoDrop 1000 spectrophotometer (Thermo Scientific).

### 3.1.6 PCR screening

Specific portions of the purified DNA were amplified via end-point PCR.

Each reaction had a final volume of 25  $\mu\text{L}$  and was consisting of:

- 16,75  $\mu\text{L}$   $\text{H}_2\text{O}$
- 2,5  $\mu\text{L}$  10X *Buffer*
- 1,5  $\mu\text{L}$   $\text{MgCl}_2$  25 mM
- 0,5  $\mu\text{L}$  dNTPs 10 mM
- 1,25  $\mu\text{L}$  *forward primer* 10  $\mu\text{M}$
- 1,25  $\mu\text{L}$  *reverse primer* 10  $\mu\text{M}$
- 0,25  $\mu\text{L}$  *Fast start Taq polymerase* 5 U/  $\mu\text{L}$  (Invitrogen)
- 1  $\mu\text{L}$  of DNA

The primers' sequence is shown in the following page, in Table 7.

Primer	Sequence
BCOR ex 15 Fw	CCATTGCAGAGGCAGAATTTTA
BCOR ex 15 Rev	CTGTACATGGTGGGTCCAGCT
BCOR intr 14 Fw	AAAGGAATGGCGATGGCC
GFP F5	CTCTCTGCTGAAACAAGCCG
GFP R5	CGGACACGCTGAACTTGTG
GFP F2	ATGCCACCTACGGCAAGCTG
GFP F3	CAAGGAGGACGGCAACATCC
GFP R2	CGCTGCCGTCCTCGATGTTG
GFP F4	AGTCCGCCCTGAGCAAAGAC
BCOR BP1 Rev	CATTTCCAAGCAACCTCAACAT
BCOR-KI_sel_FW1	AAAATCCTGCACGAGGATAAATCACG
bGH-pA-RV1	AGTTGCCAGCCATCTGTTGTTTG

**Table 7.** Primers for end-point PCR

Specific DNA sequences were amplified following the thermal protocol listed below:

1. 95°C for 5 min: enzyme activation
2. 95°C for 30 s: DNA double strand's denaturation
3. 62-66°C for 1 min: annealing between ssDNA and primers
4. 72°C for 1 min: transcripts elongation
5. 72°C for 7 min: elongation of incomplete fragments

Cycles 2 to 4 were repeated up to 40 times in order to obtain a sufficient number of copies of the specific DNA fragment.

The reactions' results were evaluated via gel electrophoresis. The gel consisted of 1,2% agarose in 100 mL of TBE buffer (Tris Borate EDTA) 0,5X to which it was added *SYBR Safe* (Life Technologies) 1X, a fluorescent dye that binds DNA double strands. To assess the DNA migration,

3  $\mu\text{L}$  of each sample and 2  $\mu\text{L}$  of bromophenol blue were loaded in an agarose gel and subjected to 120 V resistance for 30 min.

### 3.1.7 Gel extraction and Sanger sequencing

In order to sequence the PCR products, they first had to be purified. In case of multiple fragments amplified in the same sample it was necessary to select the desired bands and purify them directly from the agarose gel after the electrophoretic run.

The separated bands were cut and purified with the *QIAquick Gel Extraction Kit* (Qiagen). Every excised band was weighted and resuspended in three times the volume of QG buffer compared to the initial weight, then incubated at 50°C for 10 min, mixing the sample every 2-3 min. Once the gel was completely dissolved, 1 volume of isopropanol was added to the solution before transferring the sample into a *QIAquick* spin column (supplied with the kit) and then centrifuged at 13000 rpm for 1 min. Subsequent washes were carried out with 500  $\mu\text{L}$  of QG buffer and then 750  $\mu\text{L}$  of PE buffer, centrifuging each time for 1 min at 13000 rpm. The purified DNA was eluted by adding 30  $\mu\text{L}$  of EB buffer to the center of the column and centrifuge for 1 min at maximum velocity after 1 min incubation at room temperature.

Samples with only one band obtained from PCR were purified with a different protocol. Each sample was resuspended in 70  $\mu\text{L}$  H<sub>2</sub>O and transferred into a *MultiScreen-PCR96 filter plate* (Millipore) connected to a vacuum pump set at 100 mbar. After 10 min the plate was disconnected from the pump and the sample was resuspended by adding 30  $\mu\text{L}$  of H<sub>2</sub>O and incubating the plate on a shaker for 10 min at 400 rpm. The purified PCR fragments were in the end sequenced with the Sanger method. This methodology involves the use of sequence terminators, that are di-deoxynucleotides marked with fluorophores with a different emission wavelength for each nucleotide.

The used *Big Dye Terminator v1.1 Cycle Sequencing* kit (Applied Biosystem) requires a reaction mix composed of the following reagents:

- 2  $\mu\text{L}$  Buffer 5X  $\mu\text{L}$
- 0,32  $\mu\text{L}$  primer 10  $\mu\text{M}$  (only *Forward* or *Reverse* for each reaction)
- 1  $\mu\text{L}$  *Big Dye* 1.1
- $\text{H}_2\text{O}$  to reach the final volume of 10  $\mu\text{L}$

The primers used for the mix were the same used for PCR.

At the obtained mix was added 1  $\mu\text{L}$  of purified fragment from PCR.

The first reaction of the sequencing protocol was carried out in a thermocycler following the steps listed below:

1. 96°C for 30 sec,
2. 96°C for 10 sec,
3. 60°C for 4 min

Cycles 2 - 3 were repeated 25 times.

After the first reaction, every sample was purified by precipitating the DNA for 1 h at -20°C with a mix of 1,5  $\mu\text{L}$  of sodium acetate 3M, 30,64  $\mu\text{L}$  of ethanol 100% and 7,86  $\mu\text{L}$   $\text{H}_2\text{O}$ . After the incubation the sample was then centrifuged for 30 min at 2250 rpm at 4°C to maximize the precipitation. The supernatant was removed, and the sample was resuspended with 100  $\mu\text{L}$  of 70% ethanol before a second round of centrifugation for 15 min. The supernatant was then removed and eventual ethanol residues were made evaporate by incubating the sample at 37°C shortly.

Prior to reading the sequences with the *ABI 3730 Analyzer* (Applied Biosystem), every precipitated sample must be resuspended with 10  $\mu\text{L}$  of *Injection Buffer* (Millipore). This instrument is able to measure the fluorescence signal emitted by every sequence terminator and provide the DNA fragment's sequence.

The sequences were analyzed as electropherograms using the *Sequencer 4.1* program (Gene Codes Corporation).

### 3.2 Fluorescence In Situ Hybridization

FISH (Fluorescence In Situ Hybridization) is a technique that allows targeted analysis of chromosomal regions and permits to identify rearrangements up to 100 Kb using specific fluorescent probes.

HEK-293 cells cultured in a 25 cm<sup>2</sup> flask were detached, spinned, resuspended with pre-warmed hypotonic solution (KCl 0,075 M) (Invitrogen) and incubated at 37°C for 20 min to lysate the cells.

In the meantime, the fixing solution was prepared by mixing methanol and glacial acetic acid in a 3:1 ratio (Carlo Erba Reagenti) and cooled at 4°C.

At the end of the incubation, the cells were pre-fixed by adding 0,5 mL of fixing solution to the tube and centrifuging for 10 min at 300 g. The fixing step consisted in carefully resuspending the pellet with 10 mL of fixing solution and centrifuging again for 10 min at 300 g.

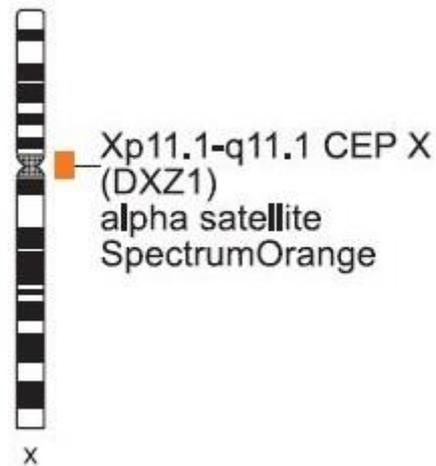
Four washes with 10 mL of fixing solution followed by 10 min centrifuge at 300 g were made to clean the sample and the cells were lastly resuspended in fixating solution.

In order to obtain an optimal spreading of the lysed cells, 2 drops of cellular solution on a clean slide and then let air-dry before being used for further analysis.

The first day of the FISH protocol is dedicated to hybridization.

The SSC 2X saline solution was incubated at 37°C for 30 min while the quality of the cells spotted on the slides were checked. Every slide was washed in the SSC 2X solution for 20 min and then dehydrated by passing it in 3 coplin jars previously cooled at -20°C and with an increasing ethanol concentration.

The probe mix had to be prepared in a dark environment in order for the probe's fluorescence not to decay. In a tube were mixed 1 µL of the probe with 2 µL H<sub>2</sub>O and 7 µL of hybridization buffer. After mixing the solution by vortexing, it was applied directly on the spotted metaphases and covered with a slide cover. The probe used was a CEP X (Xp11.1-q11.1) with an orange emission signal VYSIS (Figure 13).



**Figure 13.** *Graphic representation of the binding region of the probe used on chromosome X*

The labeled slide was incubated overnight at room temperature to allow the hybridization between the probe and the DNA. The next day is dedicated to the “high stringency” washes that are aimed to remove the excess of the probe from the sample.

Two solutions had to be prepared in advance: first an SSC solution 0,4X with 0,3% of NP40 that was incubated for 30 min at 37°C before use, and an SSC solution 2X with 0,1% of NP40 that was left at room temperature.

After removing the cover slide, the slide with the labeled sample was immersed in both solutions for 5 min in the dark, and then washed with distilled water. Lastly, 10 µL of DAPI were added to the sample as a control.

### 3.3 Transcriptome evaluation

#### 3.4.1 Total RNA extraction

Total RNA was extracted with the *RNeasy Mini Kit* (Qiagen) from 500000 cells. The cellular pellet was lysed in 350  $\mu$ L of RLT buffer and homogenized with a 0.9 Gauge syringe. After adding 350  $\mu$ L of 70% ethanol, each sample was transferred into RNeasy spin columns and centrifuged at 10000 rpm for 1 min, discarding the flow through. A solution composed of 70  $\mu$ L of RDD buffer and 10  $\mu$ L of DNase I was then added to the column and incubated at room temperature for 15 min.

Once the incubation was finished, 350  $\mu$ L of RW1 buffer were added to the sample and the column was spinned at 10000 rpm for 15 sec, discarding the flow through. A second wash was carried out with 500  $\mu$ L of RPE buffer and centrifuging the sample for 2 min. The flow through was discarded and the column was spinned again at maximum velocity for 1 min so as to eliminate the remaining traces of RPE buffer.

Each sample was eluted in 30  $\mu$ L of RNase-free water and centrifuged for 1 min. The eluted RNA was quantified with the *NanoDrop 1000 spectrophotometer* (Thermo Scientific).

#### 3.3.2 Real Time PCR

The assessment of BCOR expression levels was carried out through real time PCR.

RNA from each sample was retro-transcribed into cDNA using the *Transcriptor First Strand cDNA Synthesis kit* (Roche).

The cDNA synthesis protocol can be roughly divided in two major phases: the first consisting in the annealing between the oligo-poly(T) to the poly(A) tails of the RNAs, while the second phase is consisting in the real cDNA synthesis. In this second phase the sample was subjected to a denaturation step (65°C, 10 min), an inverse transcription step (50°C, 60 min) and the inactivation of the inverse transcriptase (85°C, 5 min).

The primers used for real time PCR are listed in Table 8.

<b>Primer</b>	<b>Sequence (5' - 3')</b>
BCOR_4840_Fw	CTCTTATGGTGCTGACCCCACC
BCOR_4957_Rev	CCACTGGCGTCATCATCATTG
GFP F5	CTCTCTGCTGAAACAAGCCG
GFP R5	CGGACACGCTGAACTTGTG
19974_ERT2	GGCCAGGCTGTTCTTCTTAG
oIMR8346_Cre	TACCGGAGATCATGCAAGC
GAPDH (201) s	CCAATATGATTCCACCCATGGC
GAPDH (318) as	CTTGATTTTGGAGGGATCTCGC
YWHAZ Fw	ACTTTGGTACATTGTGGCTTCAA
YWHAZ Rev	CCGCCAGGACAAACCAGTAT
SPP1_FW	TTTGCCTCCTAGGCATCACC
SPP1_RV	GCTTCTGAGATGGGTCAGGG
VEGFA_FW	TGAACTTTCTGCTGTCTTGGGT
VEGFA_RV	ATGTCCACCAGGGTCTCGAT
JUND_FW	CTCAAGGACGAGCCACAGA
JUND_RV	CAGCTCCGTGTTCTGACTCTT
FGF3_FW	GGGACGACTCTATGCTTCGG
FGF3_RV	CAGGGAGGACTTCTGTGTGC
ADM_FW	ATGTCGCGTCGGAGTTTCG
AND_RV	GTTGTTTCATGCTCTGGCGGTA

**Table 8.** *Primers used for real time PCR*

Each reaction was performed in a final volume of 20  $\mu\text{L}$  and was composed as follow:

- 10  $\mu\text{L}$  Sybr Green (Roche)
- 8,2  $\mu\text{L}$  H<sub>2</sub>O
- 0,4  $\mu\text{L}$  forward primer (10  $\mu\text{M}$ )
- 0,4  $\mu\text{L}$  reverse primer (10  $\mu\text{M}$ )

The cDNA was diluted and added 1  $\mu\text{L}$  for each tube. The PCR was carried out with the *FastStart Sybr Green* (Roche) polymerase through the *Light Cycler 480* (Roche).

To quantify the transcriptional level of target genes it was used the  $\Delta\Delta\text{Ct}$  method, with GAPDH and YWHAZ as housekeeping genes.

### 3.3.3 Whole Transcriptome Sequencing (WTS)

cDNA libraries were generated through the *TruSeq Stranded mRNA Library prep* (Illumina) protocol starting from 400 ng of extracted RNA. After a purification step of the poly(A)-RNA molecules with oligo-poly(T)-covered magnetic beads, the total RNA was thermically fragmented obtaining an average fragments' length of 160 bp. The fragments were annealed with random primers and underwent complementary strand synthesis obtaining double stranded cDNA (ds-cDNA). These ds-cDNA libraries were then subjected to the repair of the adhesive strands and, after being adenylated in position 3', bound to specific adapters. The modified ds-cDNA fragments were then further purified with magnetic beads, prior to an amplification step.

The quality and the dimension of the library fragments were checked by loading them on an *Agilent DNA 7500* chip and reading on a *Bioanalyzer 2100* (Agilent Technologies), while the quantification was carried out with fluorometric analysis (*QuantIT Picogreen assay*). The libraries were then diluted at 1,5 pM, inserted in flow cell and sequenced through the Illumina sequencing by synthesis technology.

### 3.3.4 Bioinformatic Analysis

The clusters images were acquired as reads that were traced back to the right sample through a base-calling process and then converted in a FASTQ file with an interconnected data-pipeline system. Based on parameters like Intensity, background and eventual non-specific signals, quality scores were assigned to the different reads. These scores permitted to exclude the reads with a low quality and high chance of misreading to improve the alignment efficiency. In cases in which only the last nucleotides of a sequence presented low quality values, or to exclude adaptors sequences, the read was trimmed, eliminating the last portion in order to obtain a final sequence with higher quality even if shorter.

For the alignment phase, the *Tophat/Bowtie* algorithm <sup>[133]</sup> permitted to align the sequences over exonic junctions and obtain the final alignment as a BAM file. The expression level of each gene was evaluated based on the concept of reading depth (or vertical coverage), that is proportional to the quantity of the different transcripts in the original sample. This calculation was made by counting the mapping reads for each gene annotated in the release v81 of the Ensemble database through the Htseq-count tool. Those data were then analyzed with R-Bioconductor.

The expression profiles of the different samples were analyzed with a non-supervised method via principal component analysis (PCA). This analysis was carried out through stats R function “prcomp”. The tool “Multiple Experiment Viewer” was used to highlight specific hierarchical clustering, exploiting the “Manhattan distance” for the metric evaluations and the “average linkage” as a clustering criterion.

The enrichment analysis to identify over-expressed pathways in the different samples was made through both *Gene Set Enrichment Analysis* (GSEA) tool and WEB-based Gene Set AnaLysis Toolkit (WEB GESTALT).

## 3.4 Cellular Assays

To evaluate the cellular growth of the different clones in normal culture conditions,  $5 \cdot 10^4$  cells from each one were plated in a 6 well plate and let grow for a maximum of 15 days. Once every 3 days the cells from one well were detached and counted manually to assess the growth rate.

The adhesion-dependent growth was evaluated both in a forced flotation condition and in invasion assays. To grow the cells in suspension,  $5 \times 10^5$  cells were plated in ultra-low attachment plates (Corning) and cultured for up to 20 days while assessing the spheroid-formation capacity.

To carry out the invasion assays, both control and mutated cells were splitted 1:1 in a new 25cm<sup>2</sup> flask 24 hours before the experiment. The transwell (Corning, with 8  $\mu$ m pores) were coated with 2,5% Matrigel in DMEM and incubated at 37°C for at least 2 hours. Just before the seeding, the excessive media was removed and replaced with DMEM 1% FBS inside the transwell insert. The cells were detached and seeded at the concentration of  $5 \times 10^4$  cells/insert, then the inserts were placed in wells containing DMEM 20% FBS plus 12,5 ng/mL FGF3 (R&D).

After 48 hours, all the cells inside the transwell were mechanically removed and the membranes were colored with Cell Stain Solution (Cell Biolab) for 15 minutes than washed in distilled water and let dry. The colored transwell inserts were put back in a dry plate and the migrated cells were counted on the microscope.

### **3.5 Protein identification**

To evaluate the effective translation of target genes or finding the molecular interactors of a specific bait protein, it was required to investigate the protein expression in the different samples.

The protein extracts were obtained as whole cell lysates, in which the cells were lysed in Ripa buffer for 30 min and then centrifuged at 14000 rpm for 25 min to eliminate the cell debris and collect the supernatant. To investigate the interactome of a specific protein it was necessary a more specific technique as the co-immuno precipitation.

#### **3.5.1 Co-Immuno Precipitation**

The co-immuno precipitation (CoIP) is a technique that permits to investigate the molecular partners of a known protein, allowing to isolate them from a protein pool obtained from cell culture or tissues. The isolation of the desired protein complex is carried out by pulling down the “bait” protein through a specific antibody. The stringency of the buffers (as in ionic force or detergent concentration) allows to co-precipitate, along with the bait protein, also the proteins that have a weak interaction with it.

The buffer A was used for the whole protocol unless indicated otherwise, and it had to be prepared the day before lysing the cells to let it cool at 4°C. The composition of buffer A is as follow:

- 10 mM Hepes pH 7.9
- 10 mM KCl
- 1.5 mM MgCl<sub>2</sub>
- 10% Glycerol
- 1mM DTT
- 1mM PMSF
- 1 tablet of Roche proteinase inhibitors (each 50 mL solution)

The cells were detached and washed three times with ice-cold PBS to eliminate traces of the media that could interfere with the protein extraction. The pellet was then resuspended in buffer A and 40 µL of 10% NOG (octyl β-D-glucopyranoside) and after an incubation on ice for 7 min it was centrifuged for 8 min at 1000 rpm. The pellet was washed three times and then resuspended in 4 mL of buffer A.

Since the genomic DNA can hinder the immunoprecipitation by covering both the proteins and antibody-coupled beads, the sample was incubated at 37°C for a total of 4 min with the addition of micrococcal nuclease (New England Biolabs). The reaction was then blocked by adding 80 µL of EGTA 1M and incubating on ice for 5 min. The sample was incubated twice for 1h at 4°C on a rotor after the addition of 750 µL KCl to increase the releasing of intracellular contents. The supernatants containing the cell lysates obtained this way were united and centrifuged at 10000 rpm for 10 min and then at 35000 rpm for 1h to remove the cell debris and the bigger aggregates respectively.

The supernatant obtained was incubated overnight with 100 µL of Anti-DDDDK-tag mAb-Magnetic Beads (MBL Science) previously washed three times in buffer A.

The next day each sample was positioned on a magnetic rack to separate the supernatant from the beads bind to the bait protein complex, which was washed three times with 1 mL of buffer A plus of 180 µL KCl 1M.

The proteins bound to the washed beads were then evaluated in western blot analysis.

### 3.5.2 Western blot analysis

The evaluation of the presence of specific proteins in a cellular lysate or in samples from selective methods as the Co-IP was made through western blot.

For each protein sample it was prepared a mix 3:1 with Laemmli buffer (Bio-Rad) and 50mM  $\beta$ -MercaptoEthanol, and incubated 5 min at 98°C to denature the proteins.

The unfolded proteins were separated on a polyacrylamide gel (SDS-PAGE) subjected to a constant resistance of 40V. The migration takes place in two phases with different pH and ionic force, named *stacking* and *separating gel*. The stacking gel includes the wells in which the samples are loaded and is responsible of the alignment of the proteins before the separating gel. The second can be made at different acrylamide's concentration depending on the target protein's dimension. In this case it was used a 10% gel to resolve proteins with molecular weight higher than 60 KDa, while a 6% gel was used for smaller proteins.

After the electrophoretic run, the proteins contained in the gel are transferred to a polyvinylidene fluoride (PVDF) membrane with Trans-Blot Turbo Transfer system (Bio-Rad).

The membrane containing the transferred proteins was cut and washed in TBS 0,1% Tween 20 for 10 min, then incubated with a blocking solution (TBS-T, 5% Milk) in constant tilting at 4°C for 1h. The specific primary antibody was added directly to the blocking solution (working dilution in Table 9) and then the membrane was incubated overnight in constant tilt at 4°C.

<b>Antibody</b>	<b>Working dilution</b>
M2 Anti-Flag (Thermo Fisher)	1:1000
Anti-BCOR (Cell Signaling)	1:1000
Anti-RING1B (Cell Signaling)	1:1000
Anti-PCGF1 (Bardwell lab)	1:1000
Anti-RYBP (Cell Signaling)	1:1000
Anti-SKP1 (Cell Signaling)	1:1000
Anti-Cre recombinase (Thermo Fisher)	1:5000
Anti-H3 (Cell Signaling)	1:3000

**Table 9.** *Primary antibodies used for Western Blot analysis*

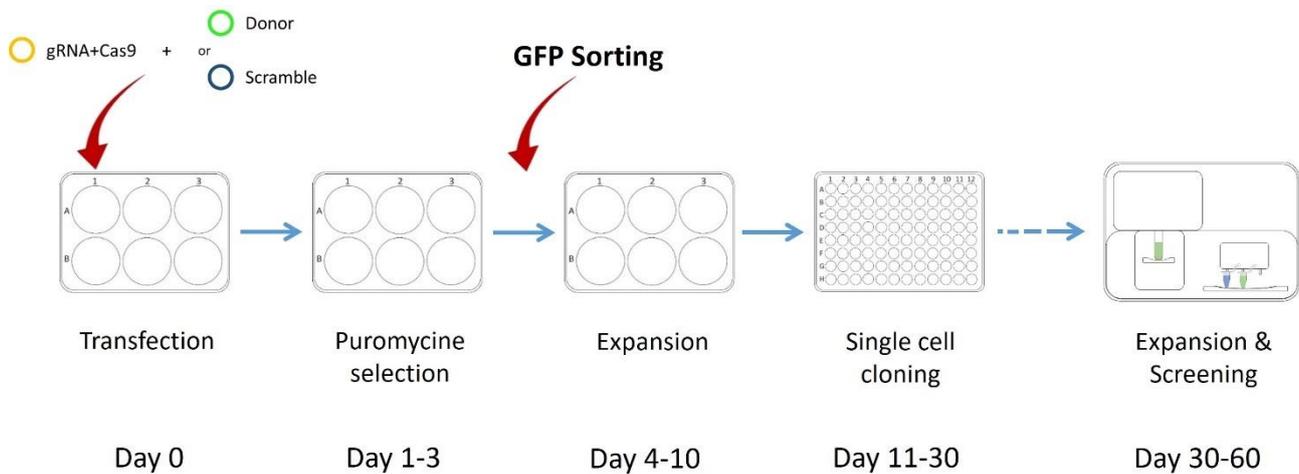
The next day the solution containing the primary antibody was removed and the membrane was washed three times with TBS-T before incubating it with the appropriate secondary antibody with a dilution of 1:30000 in TBS-T 5% milk for 1 h.

After this incubation the membrane was washed again for three times in TBS-T and detected at the chemidoc (Bio-Rad) by using SuperSignal west Atto ultimate sensitivity substrate (Thermo Fisher).

## 4. RESULTS

### 4.1 Gene editing in HEK-293

In order to obtain a preliminary study model for CCSK presenting alterations at the level of the BCOR gene, a gene editing strategy was adopted with the aim of introducing a tandem duplication (ITD) of 90 bp exactly before the stop codon in the last exon of BCOR in HEK-293 cells (Fig.14).



**Fig 14.** Schematic representation of the workflow to obtain recombinant clones

The chosen approach involved the use of the CRISPR/Cas9 system with two types of vectors: a Guide vector coding for puromycin resistance, for the Cas9 nuclease together with a guide RNA which induces site-specific cutting of the enzyme, and a Donor vector containing the sequence to be inserted into the genome of the target cells with upstream and downstream homology sequences of 800 bp in order to facilitate the alignment and integration in the site of interest (Fig.15).

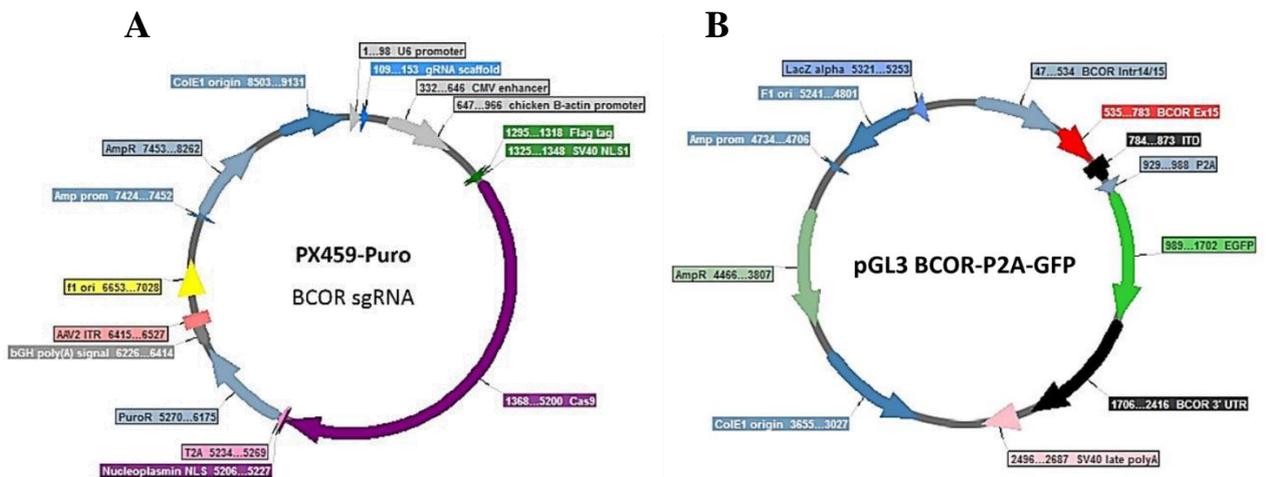


Fig 15. Schematic representation of the (A) Guide and (B) Donor plasmids.

In order to facilitate the selection of effective recombinant clones, the Donor vector was designed to link the last BCOR exon to the polycistronic linker P2A upstream of the GFP gene (Fig. 15 B). The P2A-GFP sequence was inserted between the end of BCOR's last exon and its 3' UTR sequence, eliminating BCOR stop codon in order to allow the construct's transcription as a whole.

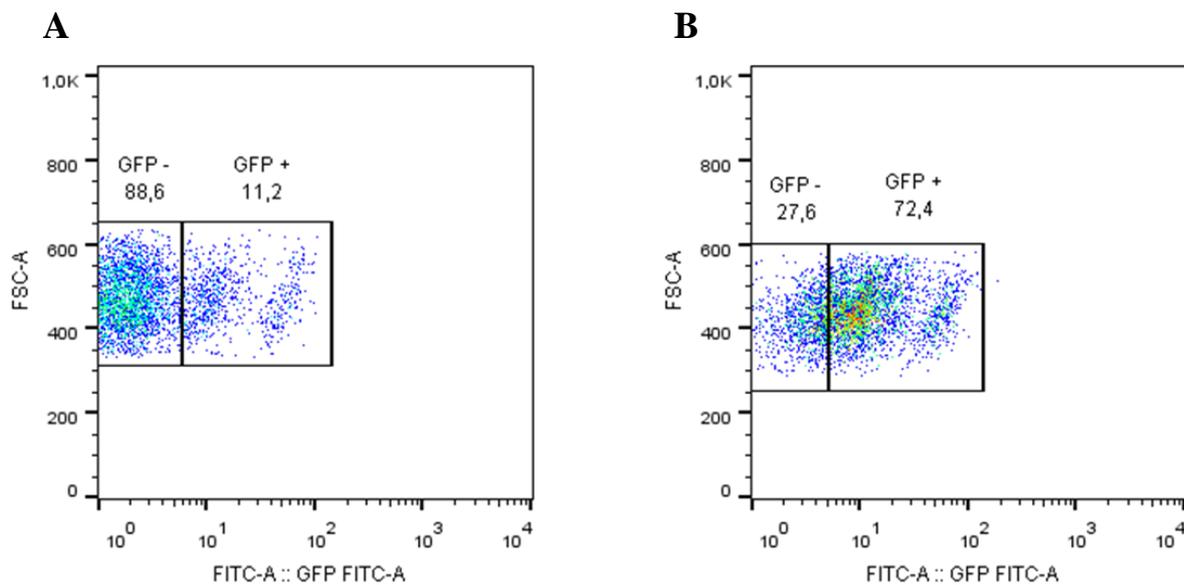
The 14th intron of BCOR, together with exon 15 bearing the 90 bp ITD and the 3'UTR sequence were amplified from a previously characterized CCSK sample [52]. The different amplicons were linked together via fusion PCR, by using primers with overlapping overhangs. The Scramble plasmid, carrying a non-specific guide RNA sequence, was used as control.

After 24 h from transfection, effectively transfected cells were selected by Puromycin treatment. Resistance to the antibiotic was provided by the guide plasmid which, being transiently transfected, would have been lost within the first 72 hours. The surviving cells belonging to each experimental group were then grown and selected by means of flow cytometry cell sorting for GFP expression. After sorting, the cells were cloned in single-cell in 96-well plates and each clone was amplified to be used for subsequent analyses.

A similar cloning was also performed for the cells treated with Scramble vector, kept as bulk from the first passages.

#### 4.1.1 Clones screening by flow cytometry analysis

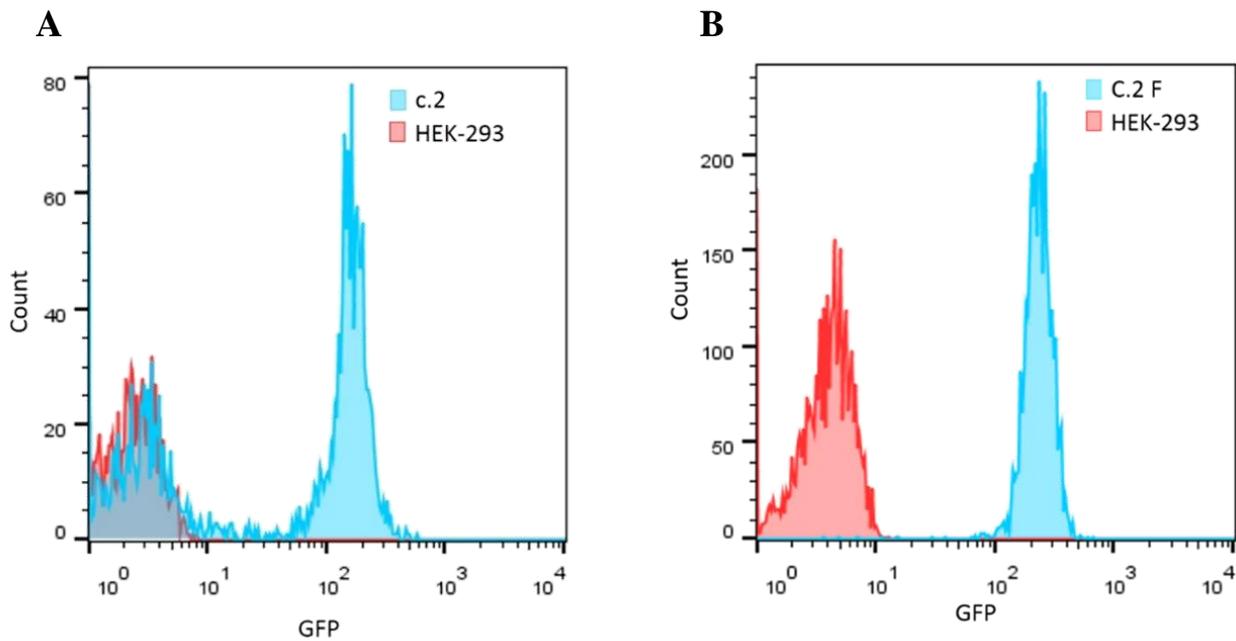
After the first selection with Puromycin, the surviving cells were expanded and then analyzed by flow cytometry to evaluate GFP expression's enrichment in the sorted bulk cultures (Fig. 16 B) compared to the non-sorted cells (Fig. 16 A)



**Fig 16.** Evaluation of the GFP<sup>+</sup> cells enrichment after cell sorting. (A) non-sorted sample; (B) sorted sample.

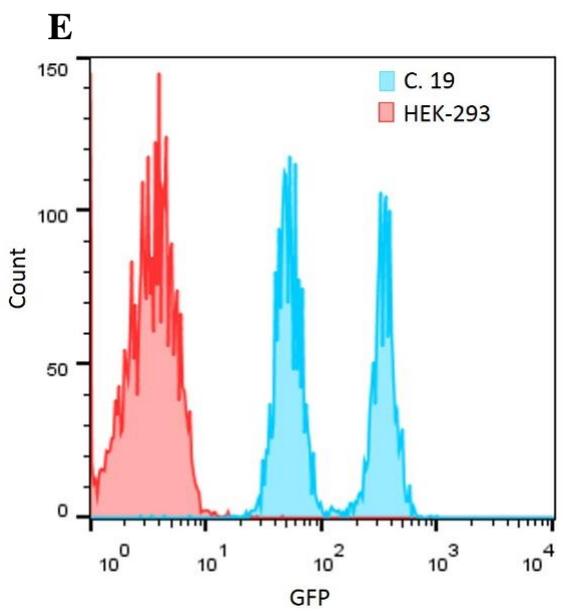
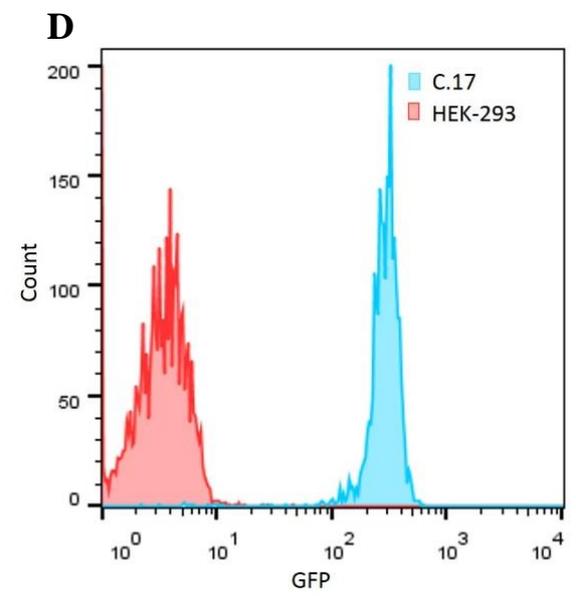
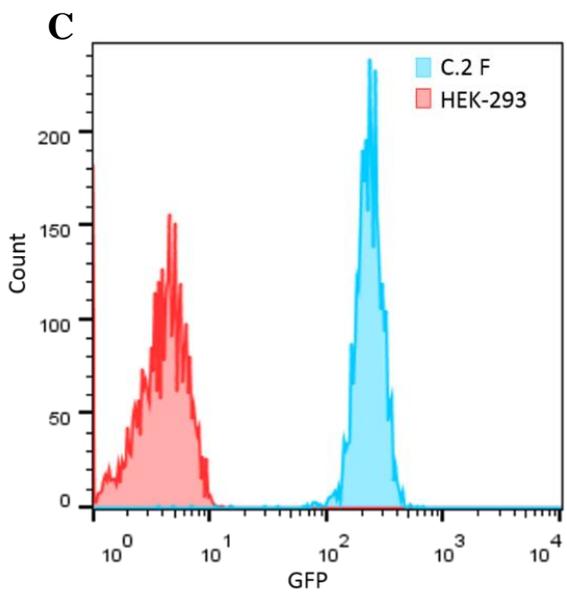
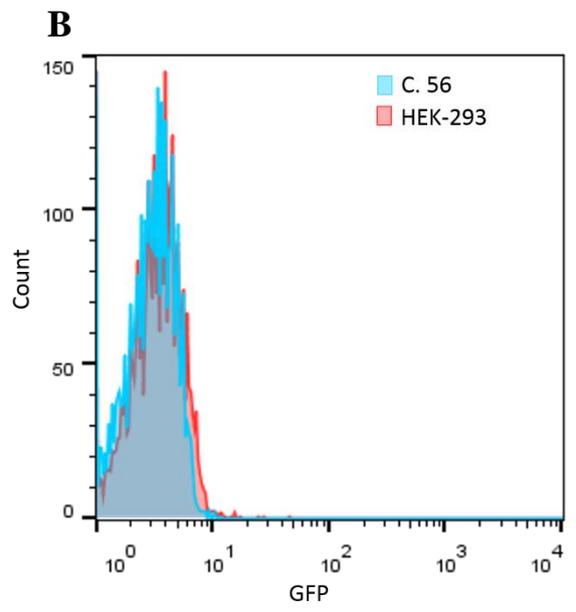
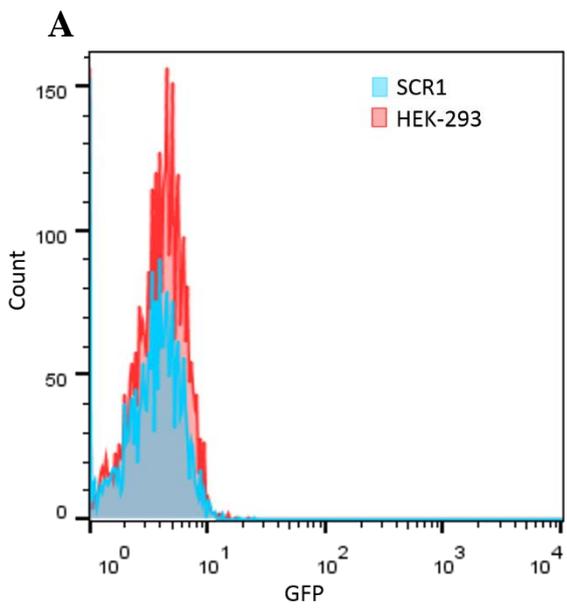
From the comparison between sorted and non-sorted populations, is clear that the sorting process permitted a significative, although not complete, enrichment in the GFP<sup>+</sup> population.

A total of 99 clones were obtained from the single-cell cloning process (numbered from c.1 to c.99). To screen for the effective insertion of the construct, each clone was evaluated for its GFP expression, finding out a total of 12 GFP<sup>-</sup> clones, 83 GFP<sup>+</sup> clones and 4 clones with both GFP<sup>+</sup> and GFP<sup>-</sup> populations. One of the latter undergone a sub-cloning, obtaining GFP<sup>+</sup> clones as reported in Fig 17.



**Fig 17.** GFP expression analysis on the obtained clones via flow cytometry. Analysis before (A) and after (B) sub-cloning and selection of the GFP<sup>+</sup> colonies

Amongst the GFP<sup>+</sup> clones identified, three clones with high GFP expression were selected for further characterizations, along with a clone obtained from the transfection with the Scramble vector (SCR1). As a negative control for both GFP expression and BCOR alterations it was used a previously obtained GFP<sup>-</sup> clone that displayed the complete loss of the last exon of BCOR (C.56). This results in a truncated protein missing of a complete PUF<sup>2</sup> domain and thus considered as a functional K.O..

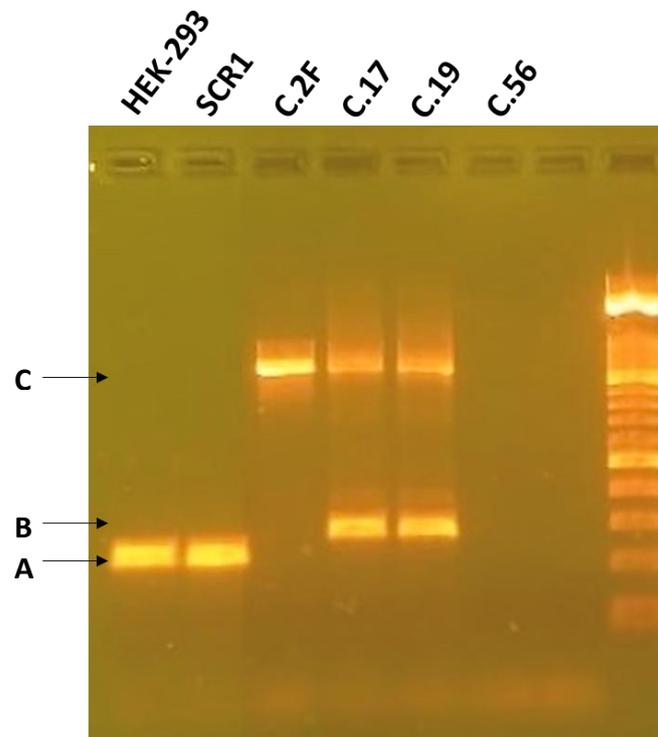


**Fig 18 -on the previous page-** . *GFP expression of the clones selected for further analysis. (A, B) Scramble and C.56 clones were used as GFP<sup>-</sup> controls. (C, D, E) Recombinant clones selected for GFP expression.*

Amongst the clones selected, C.2F and C.17 show high GFP expression's levels, while C.19 shows a heterogeneous population with two fluorescence peaks with different intensities. SCR1 and C.56 fluorescence signals overlapped completely with the parental cell line's.

#### 4.1.2 Identification of editing events

To verify the correct knock-in of the construct in the different clones, a control PCR was performed using primers designed to pair across the predicted insertion site. In the wild type controls and in the cells in which no recombination occurred, a single band of 221 bp is obtained, while in case of complete recombination a single band of greater size than in the WT (1099 bp) was observed, due to the insertion of the 90 bp ITD tailed by P2A and GFP (Fig 19). In the C.17 and C.19 clones a partial insertion was seen, that led to a 273 bp band.



**Fig 19.** *Agarose gel run of PCR samples from the selected clones. (A) bands corresponding to BCOR wild type. (B, C) Amplicons reporting partial (B) or correct insertions (C).*

Sanger sequencing of the bands extracted from the agarose gel allowed to identify the inserted fragments sequence.

The analysis of the electropherograms highlighted both the presence of a 13 bp deletion and the insertion of a 70 bp portion of the GFP sequence within the lowest band of clones C.17 and C.19, which makes them composite heterozygous. The C.2F clone is instead homozygote for the construct, while the C.56 clone had already been shown to be missing the last portion of BCOR.

Clone	BCOR allelic recombination
C.2F	c.[5136_5225dup; 5265_5268delins792]
C.17	c.[5136_5225dup; 5265_5268delins792]; c.5255_5268delins70
C.19	c.[5136_5225dup; 5265_5268delins792]; c.5255_5268delins70
C.56	Exon 15 deletion
SCR1	- none -

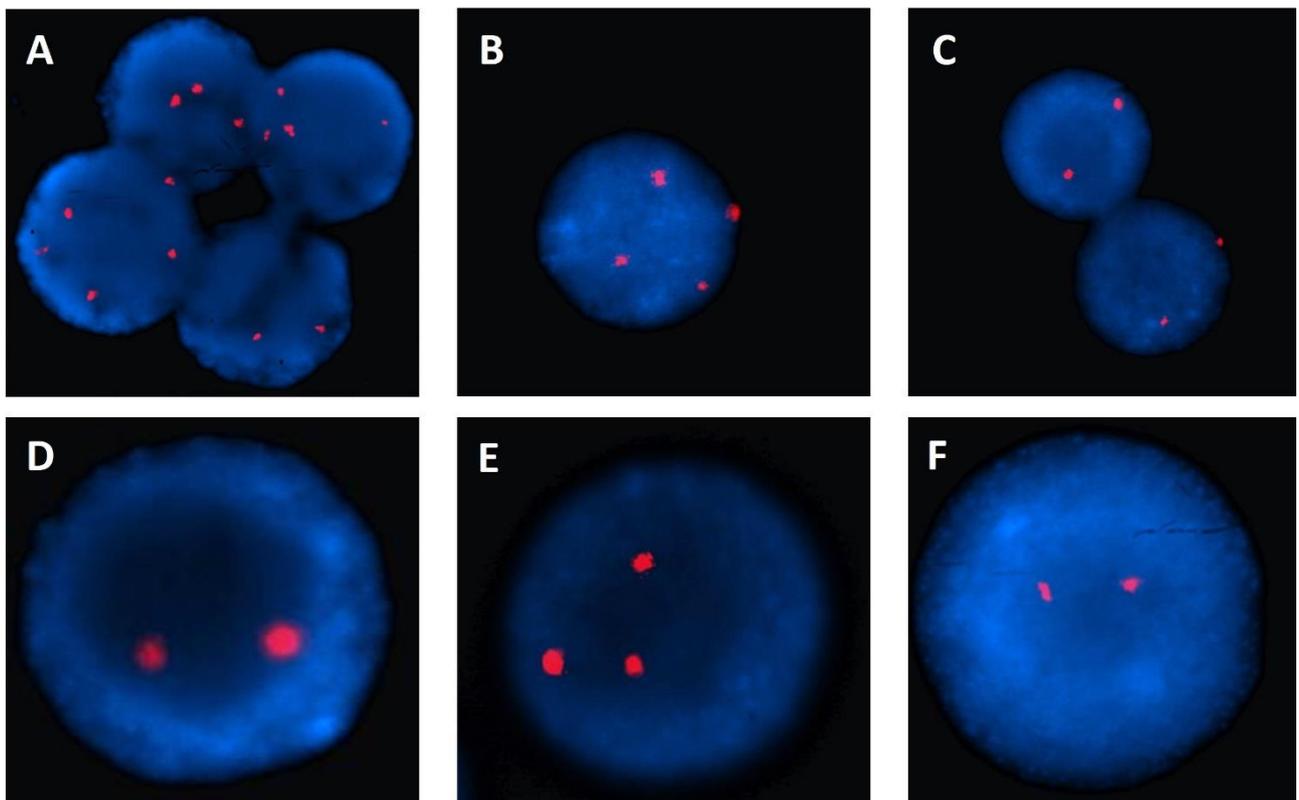
**Table 10.** Schematic summary of the allelic recombinations in the analyzed clones.

### 4.1.3 Clones' characterization

Fluorescence in situ hybridization allowed to define the status of the clones at the chromosomal level, as the locus in which BCOR is located is on the short arm of the X chromosome.

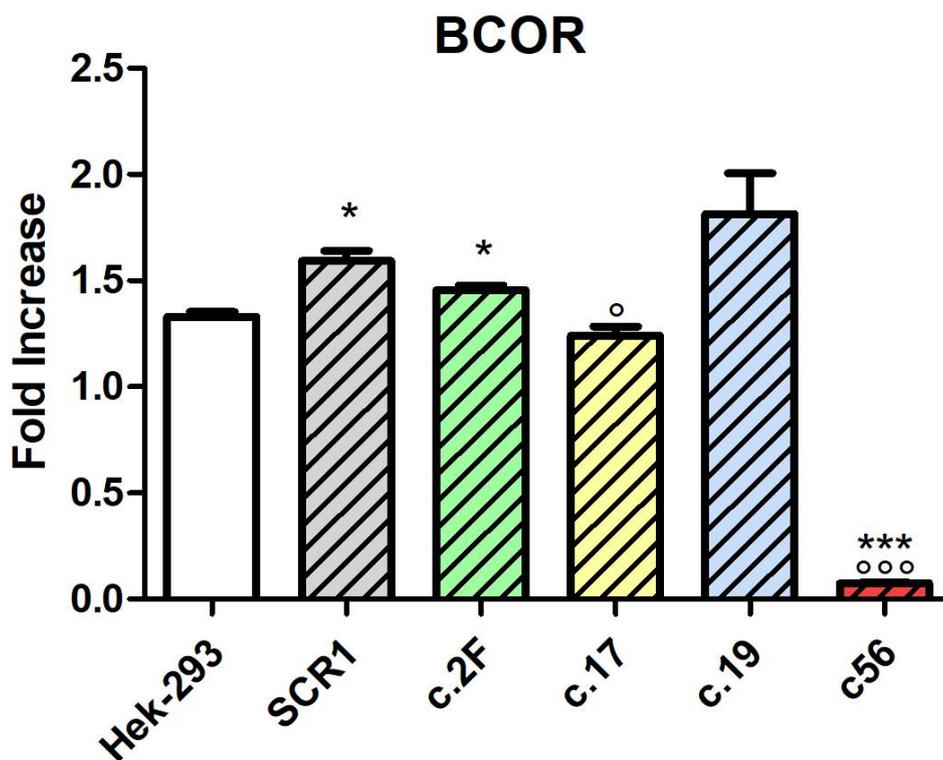
In the literature, HEK-293 cells present a state of aneuploidy, with multiple copies of the X chromosome. This characteristic is also highlighted by the FISH analysis reported in Fig 20, where an average of three X chromosomes was observed in 80% of the cells. The Scramble clone has the most X chromosomes, with four copies in 86% of cases.

In the case of clone C.56, the use of the CEP X (Xp11.1-Xq11.1) probe highlighted the presence in almost all cells (98%) of two X chromosomes. Regarding the recombinant clones, however, the homozygous clone (C.2F) was diploid in 97% of cases, while for heterozygotes C.17 and C.19 respectively 75% of XXX and 65% of XX were observed.



**Fig. 20.** Quantification of the number of X chromosomes in FISH. (A) HEK-293, (B) SCRI, (C) C.56, (D) C.2F, (E) C.17 and (F) C.19

In order to verify if the different numbers of X chromosomes in the different clones was leading to different expression of BCOR, a Real Time PCR analysis was carried out with primers designed between the 12<sup>th</sup> and the 13<sup>th</sup> exons of BCOR (not involved in the recombination process). Despite an overall similar BCOR expression amongst the clones, the C.2F clone revealed a slight but significant increase of expression compared to the parental HEK-293. The SCR1 clone, despite being used as a control, showed a significant increase in BCOR expression compared to the parental cells, most probably due to the presence of four X chromosomes. Noteworthy is the almost absent expression in the C.56 clone.



**Fig 21.** Real Time PCR evaluation of BCOR expression. \* relative to HEK-293; ° relative to C.2F

In order to evaluate the oncogenic hit given by the inserted ITD, the growth rate of the different clones was tested, revealing no significant variations in normal culture conditions over 12 days.

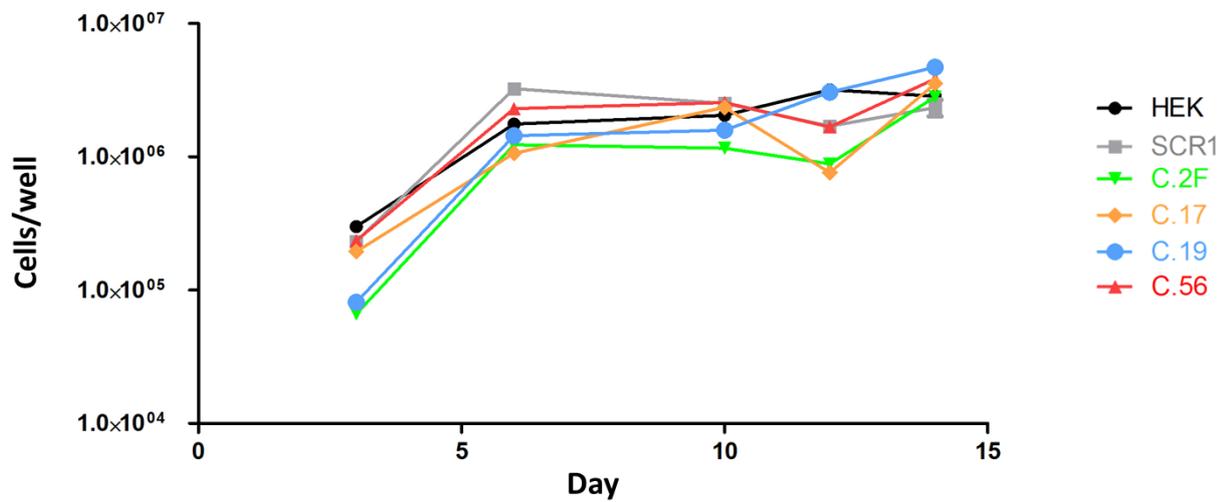
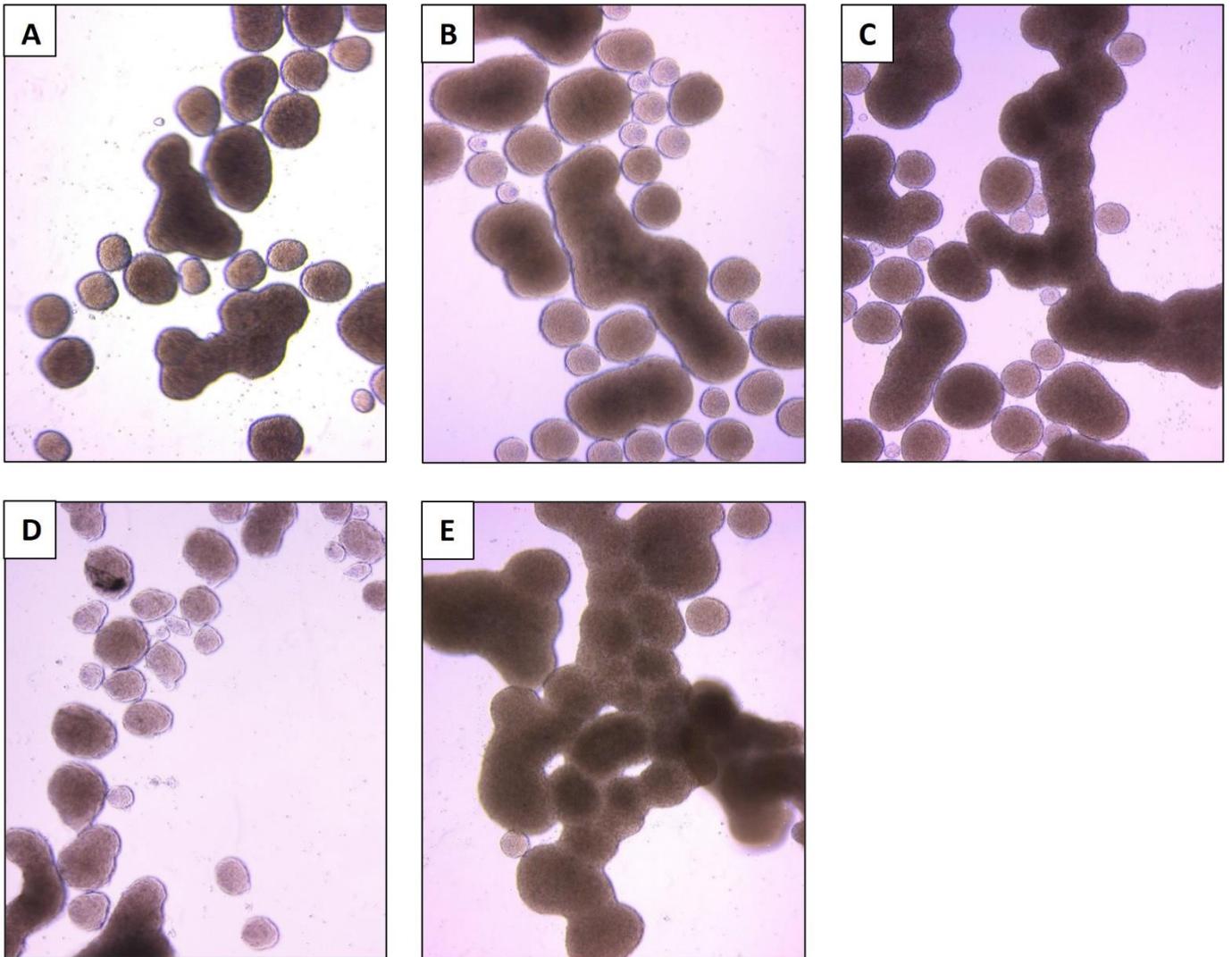


Fig. 22. Growth rate evaluation over 12 days coltures.

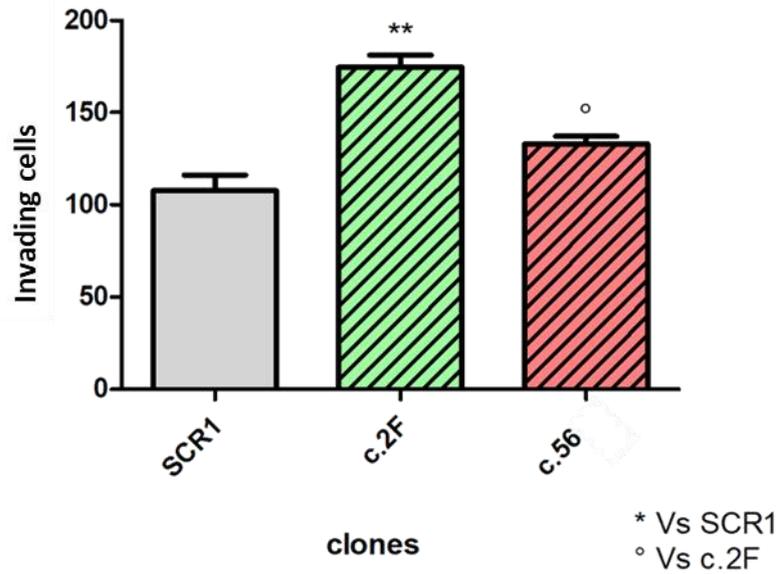
Adhesion-dependent growth was evaluated by growing the different clones in ultra-low adhesion plates for 20 days. Contrary to what was seen in the control, the C.2F clone demonstrated a severe growth impairment, generating principally small and slow-growing spheroids and presenting a higher amount of dead cells, while both C.17 and C.19 shown an intermediate growth capacity. The C.56 clone, on the contrary, demonstrated a high growth capacity.



**Fig. 23.** *Evaluation of adhesion-dependent growth. (A) HEK-293 as a control; (B, C) heterozygotes clones C.17 and C.19; (D) homozygote clone C.2F; (E) C.56 clone.*

The large difference in terms of adhesion-dependent growth rate between control, homozygote clone and the functional K.O. (Fig.23 A, D and E respectively) was further evaluated through invasion assay.

As shown in Fig, 24, the C.2F clone showed a significantly increased number of invading cells compared to both the control and the C.56 clone, meaning of an increased invasive capacity.

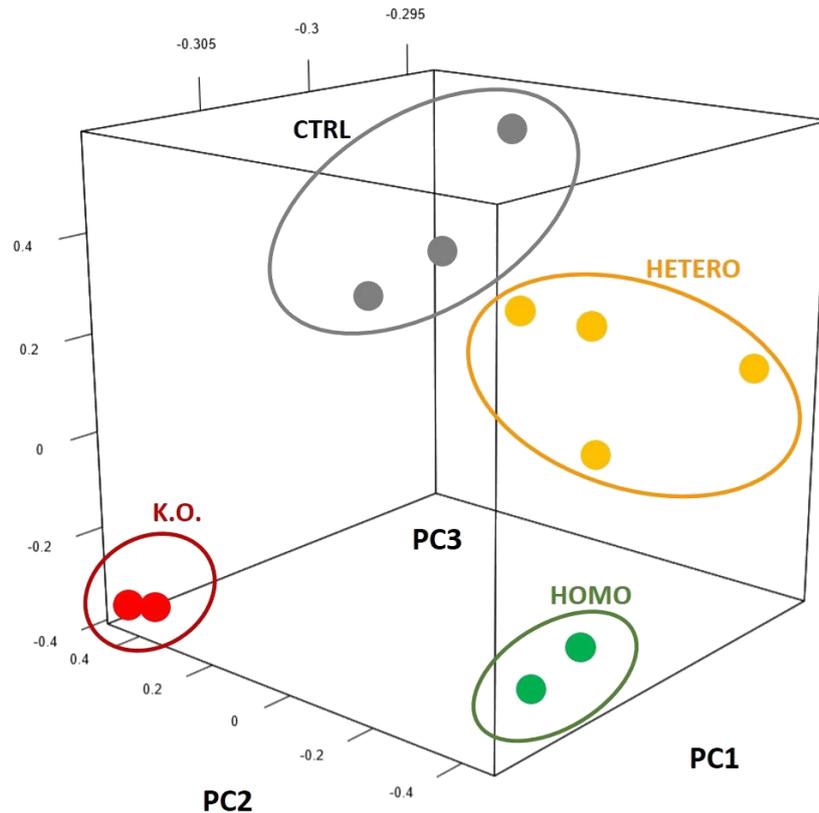


**Fig. 24.** Comparison of the clones' invasive capacity

#### 4.1.4 Gene expression profiling

WTS data from biological duplicates of each clone were analyzed to quantify gene expression, using HEK-293 biological duplicates and SCR1 as reference group. The aim of this analysis was to identify the gene sets that characterize the ITD clones and compare them with what is known in literature about CCSK.

The global profile of the transcriptome was investigated by principal component analysis (PCA), which allowed to group the different samples' gene expression through a graphical correlation (Fig. 25).



**Fig.25.** *Graphic representation of the first three principal components of the global expression profiles. The homozygous clone (green) segregates from CTRL lines (grey) along the 2nd and 3rd component. The K.O. clone (red) segregates independently and far from the homozygous one.*

The principal component analysis revealed how biological duplicates segregate in the same way. A clear distinction along the first component separates the wild type controls from the homozygote clone (C.2F) and the K.O. (C.56), while the heterozygote clones (C.17 and C.19) colocalize midway. It is also noteworthy a further separation along the second component between C.2F and C.56.

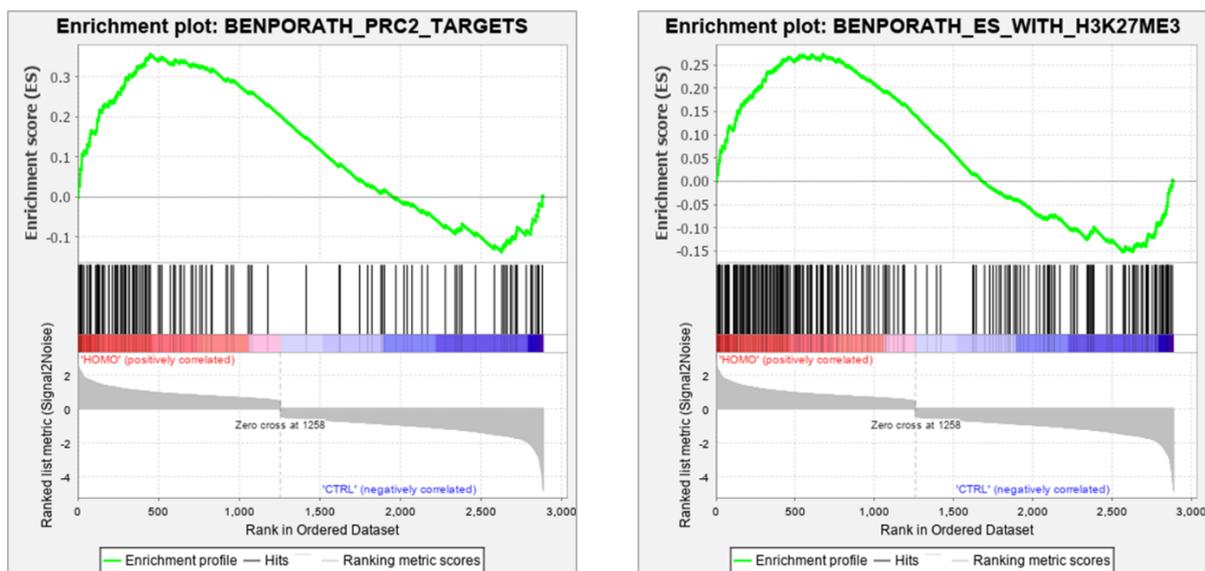
The supervised analysis identified around 1400 differentially expressed transcripts with a p-value lower than 0,05 between the recombinant clones and the controls.

In order to recognize, in the context of a broad molecular signature, the most significant pathways characteristic of the BCOR-ITD clones, an enrichment analysis was carried out with the GSEA tool, using both pathways obtained from public databases and gene-sets derived from the literature.



Among the most significant pathways were those of the PRC2 complex's methyl transferase activity targets, in which methyl groups are added to Lys27 in histone H3 (Fig. 26).

The enrichment of these pathways, especially seen in the homozygote clone (Fig. 27), means the loss or impairment of the function of the PRC2 complex, which acts as a transcriptional repressor. This result is in line with the known crosstalk between the activity of the PRC1 complex, both canonical and non-canonical, such as the one in which BCOR participates (PRC1.1), with the methyl-transferase activity of the PRC2 complex, in which they participate EZH2, EED and SUZ12.



**Fig. 27.** *Enrichment Plots relating to genes modulated by PRC2 activity in C2F compared to the controls*

The analysis of the enriched pathways also identified an enrichment of genes involved in AKT and PI3K pathways, as well as WNT pathway.

Moreover, the enrichment of genes involved in cell adhesion and interaction with the extracellular matrix (ECM), especially in the homozygote clone, are supporting evidence to the adhesion-dependent growth of the C2F clone.

## 4.2 Characterization of metastasis-driving factors

### 4.2.1 Identification of a peculiar signature for metastatic CCSKs

The study series consisted of 8 patients affected by Clear cell sarcoma of the kidney (CCSK) and enrolled in the TW-2003 AIEOP protocol in a period between 2003 and 2005 [134].

The diagnosis was made by histological analysis of the operative specimen or surgical biopsy and confirmed by centralized review at the National Reference Center for Pediatric Renal Tumors, at the IRCCS and the Istituto Nazionale dei Tumori.

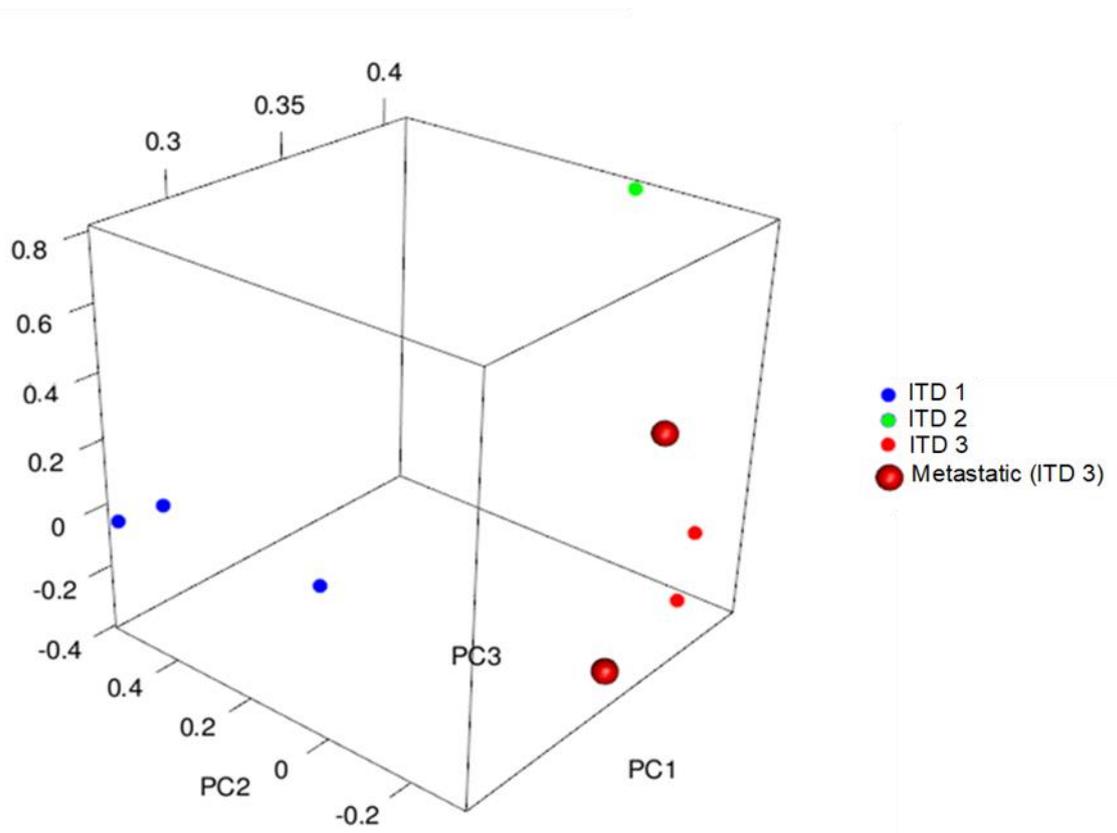
ID	Sex	Age (months)	Side	Stage	Metastatic site	Relapse	BCOR molecular analysis
CCSK1	M	37	L	I	-	N	c.5171_5266dup
CCSK2	F	20	L	III	-	N	c.5163_5225dup
CCSK3	M	27	R	IV	Femur	Iliac wing	c.5171_5266dup
CCSK4	M	14	R	I	-	N	c.5163_5225dup
CCSK5	M	21	L	IV	Lung	N	c.5171_5266dup
CCSK6	M	27	L	I	-	N	c.5171_5266dup
CCSK7	F	17	L	II	-	N	c.5163_5225dup
CCSK8	F	25	R	II	-	N	c.5099_5212dup

**Table 11.** Schematic representation of the analyzed cases.

From all patients, a frozen tissue sample of the primary tumor was preserved at the time of surgery and was used for both the molecular analysis of BCOR and the study of the gene expression profile by Whole Transcriptome Sequencing (WTS). For all the CCSKs used, the presence of the Internal Tandem Duplication in the last exon of BCOR was previously confirmed by PCR and Sanger sequencing. Three different ITDs were identified in the eight patients, and more specifically c.5136\_5225dup (ITD1), c.5099\_5212dup (ITD2) and c.5171\_5266dup (ITD3) [52].

Whole transcription sequence data from the eight CCSK cases were analyzed and compared in order to identify a specific signature for the metastatic cases.

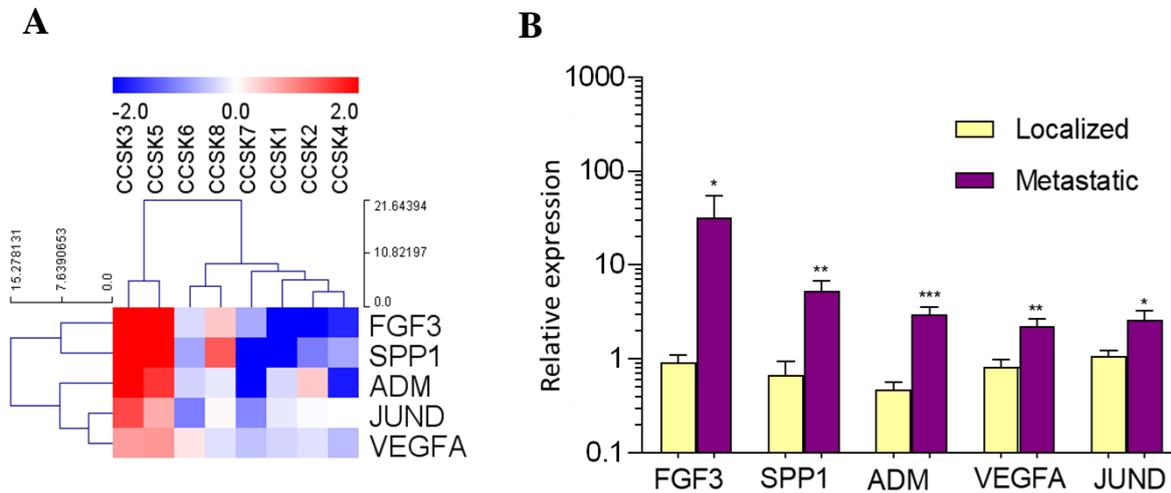
The unsupervised analysis of whole transcriptome sequencing (WTS) data on fresh frozen samples did not reveal any clear segregation of the global expression profiles of the cases with metastatic-onset and the localized ones (Fig. 28). Conversely, an evident clustering of samples based on the type of ITD carried was found, with a clear segregation between the group of three CCSKs characterized by ITD-1 and that of four CCSKs characterized by ITD-3 on the second principal component.



**Fig. 28.** *Principal component analysis (PCA) of the global expression profiles of the different CCSK cases.*

The supervised analysis of the expression profile of metastatic cases compared to localized ones identified 783 genes differentially expressed with a p-value <0.05 and 156 with a p-value <0.01. The pathway enrichment analysis identified several protein-coding genes associated with "MAPK signaling pathway" significantly up-regulated in metastatic cases (11 genes; adj-p<0.0001). The

functional analysis of the molecular signature of metastatic cases highlighted the statistically significant over-expression of 5 genes: FGF3 ( $p=0.0006$ ), VEGFA ( $p=0.0009$ ), SPP1 ( $p=0.0003$ ), ADM ( $p=0.0009$ ) and JUND ( $p=0.004$ ) (Fig. 29 A).



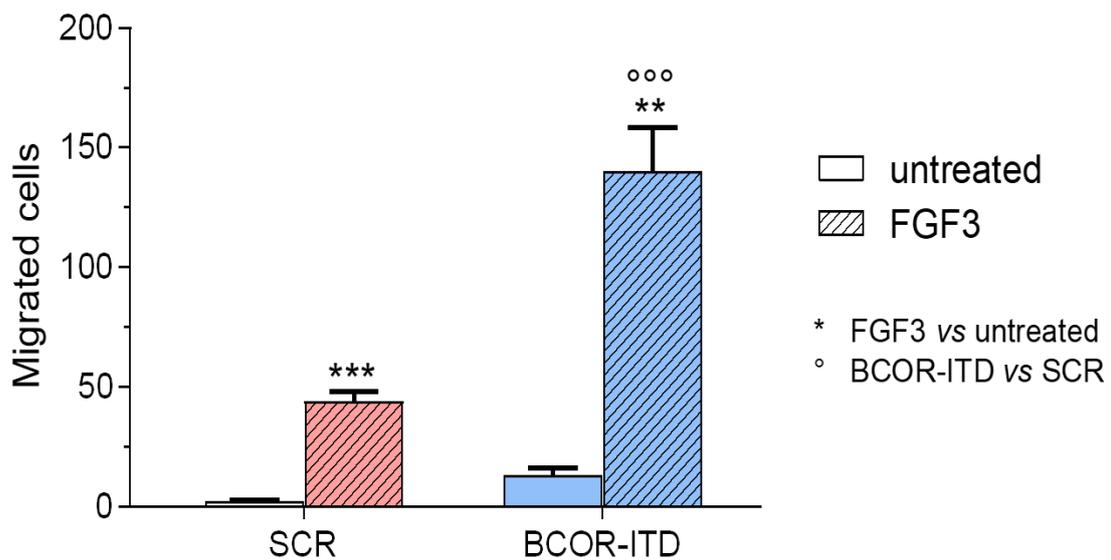
**Fig. 29.** Representations of the differentially expressed genes between metastatic and localized tumor samples. (A) Heatmap representation; (B) Validation of the differentially expressed genes by real time PCR.

Data from WTS were validated by quantitative PCR, revealing a 20-fold expression of FGF3 in the metastatic samples compared to the localized ones. This result suggested that FGF3 expression could be casually connected to the acquisition of a more aggressive phenotype.

#### 4.2.2 In-vitro validation of the metastatic drive

The functional role of FGF3 was investigated in-vitro by analyzing the induced cellular invasion of the before-mentioned homozygote clone obtained in HEK-293 cells, as a model reproducing the oncogenic hit of CCSK.

As reported in Fig. 30, migration assays shown a specific and significant increase in cell migration over the untreated cell lines, with a significant higher induction of cell migration on the BCOR-ITD with respect to the SCR1 clone.



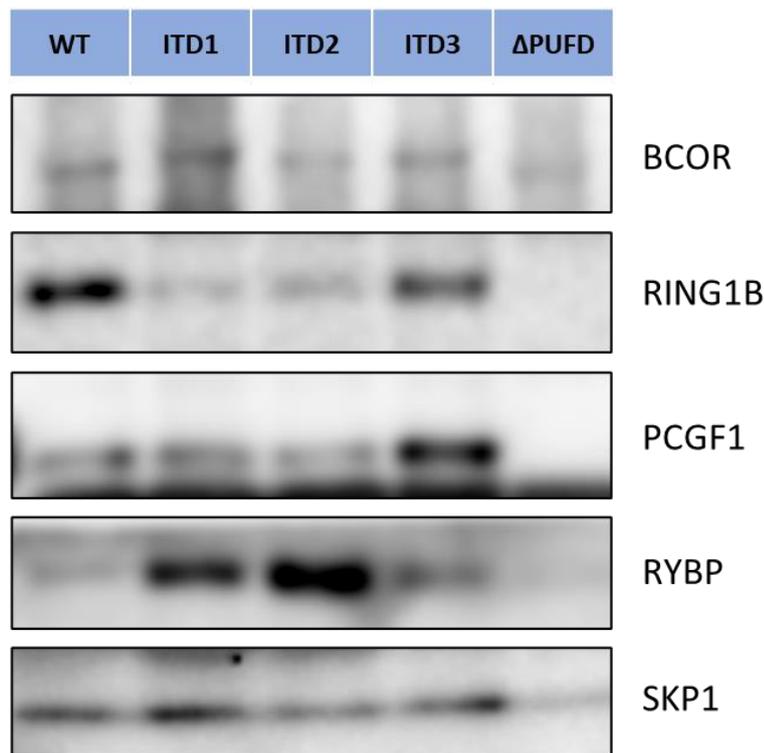
**Fig. 30.** Evaluation of the number of invading cells after treatment with FGF3

### 4.3 PRC1.1 reconstruction

In order to evaluate the PRC1.1 complex formation in presence of BCOR-ITD, the core components of the complex were transiently overexpressed in HEK-293 FT along with different BCOR variants.

The overexpression plasmids carrying the BCOR variants were previously obtained by inserting the ITD1 (114 bp), ITD2 (90 bp) or ITD3 (96 bp) according with the previous findings in Astolfi et al. (2015). The  $\Delta$ PUFD BCOR variant was obtained by deleting the portion related to the PUFD domain. This variant was used as a negative control since it lacked the binding site with several core components of the complex.

The isolation of the reconstructed complexes was carried out via Co-Immunoprecipitation by targeting the Flag-tag located at the N-Terminal end of each BCOR variant.



**Fig. 31.** *Western Blot analysis of the CoIP samples.*

The different weight of the BCOR variants, as shown in Fig. 31, reflected the insertions (in ITD variants) and deletions (in the  $\Delta$ PUFD sample) of aminoacids to the wild type sequence.

The different amounts of precipitated cofactors for each variant were normalized on BCOR signal and reported in Fig.32.

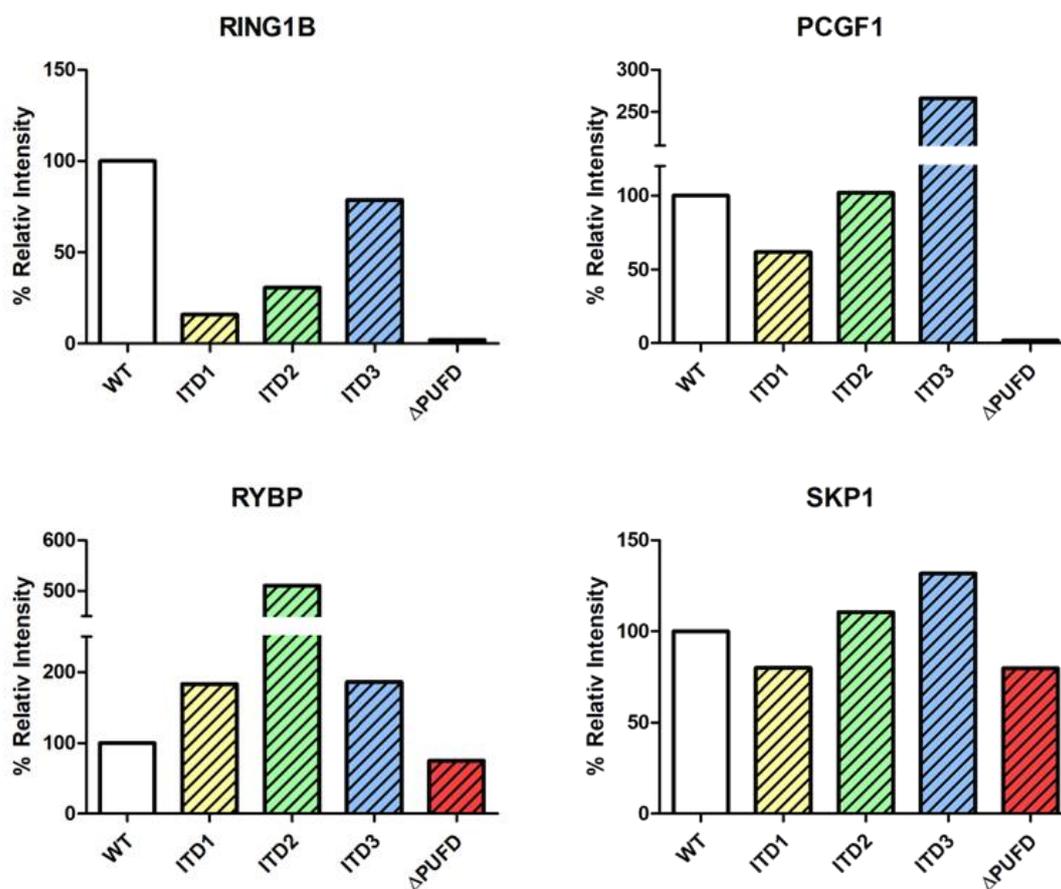


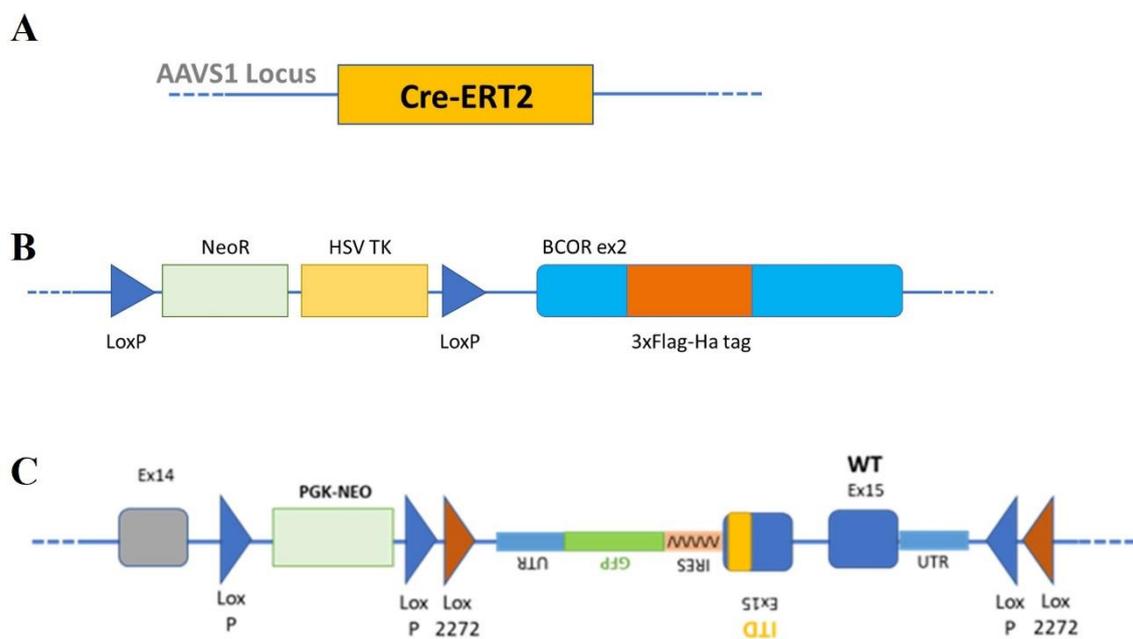
Fig. 32. Relative intensity of the signals relative to the precipitated cofactors.

The relative quantification of the precipitated co-factors revealed that every BCOR variant possessed a different binding capacity, especially with regard to those cofactors that directly interact with the BCOR protein, such as PCGF1. The  $\Delta$ PUFD BCOR variant, depleted of the PUFD domain, is unable to bind both PCGF1 and one of the enzymatic cores of the complex, RING1B, which exhibits ubiquitin ligase activity.

## 4.4 Inducible model design

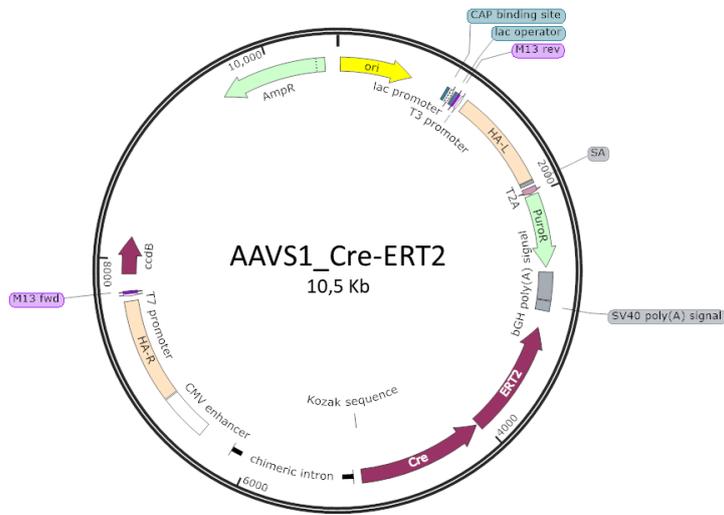
In order to obtain a cellular model for the study of BCOR-ITD functions in pluripotency maintenance and correct differentiation, it was designed an inducible model in H1 cells.

The multi-step design for the insertion of different constructs (as schematized in Fig. 33), and still ongoing, required recombination processes through both TALENs or CRISPR/Cas9 system.



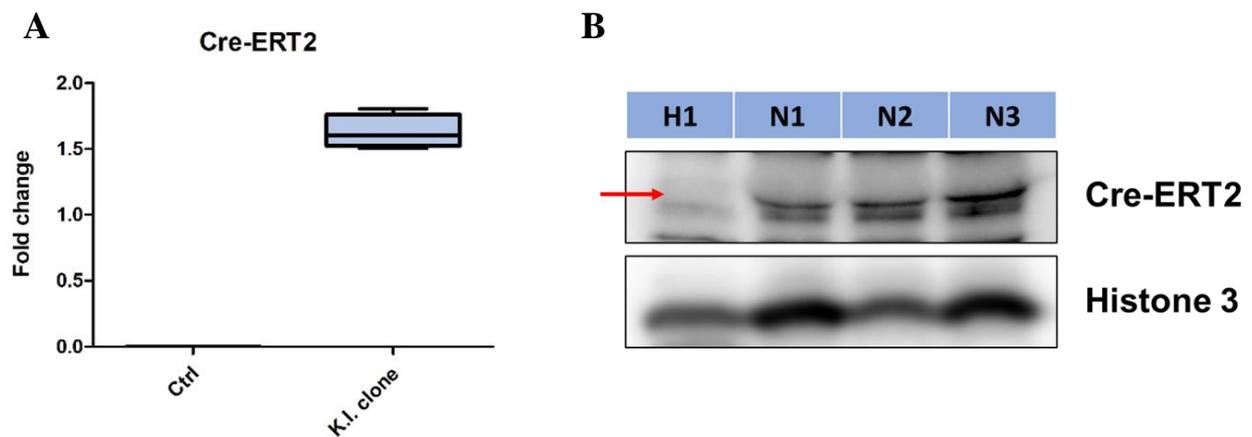
**Fig. 33.** Schematic representation of the multi-step design. (a) First, insertion of the Cre-ERT2 chimeric gene in the AAVS1 locus, followed by (b) insertion of a 3XFlag - HA tag at the second exon of BCOR. (c) “Flippable” BCOR exon 15.

The insertion of a tamoxifen-inducible CRE-ERT2 chimeric gene in the AAVS1 locus was executed by co-transfection of AAVS1-TALEN-L and AAVS1-TALEN-R plasmids (Addgene) with a donor vector carrying the chimeric gene and the puromycin resistance cassette downstream of the CMW enhancer.



**Fig.34.** Schematic representation of the Cre-ERT2 vector

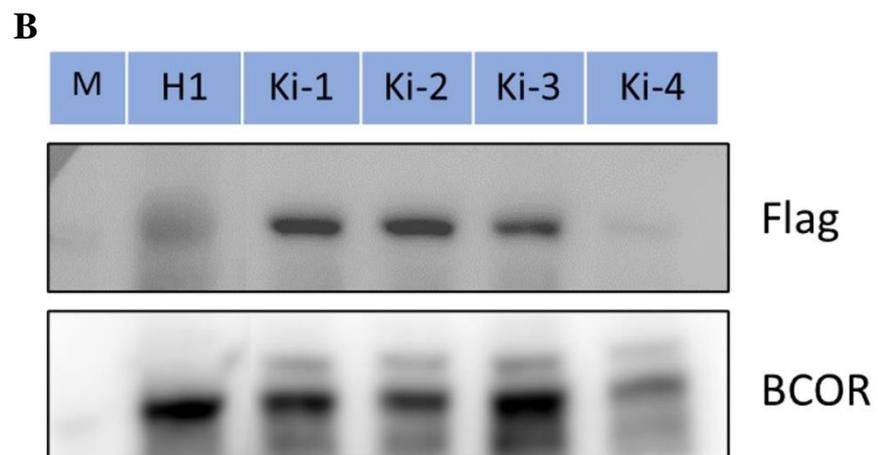
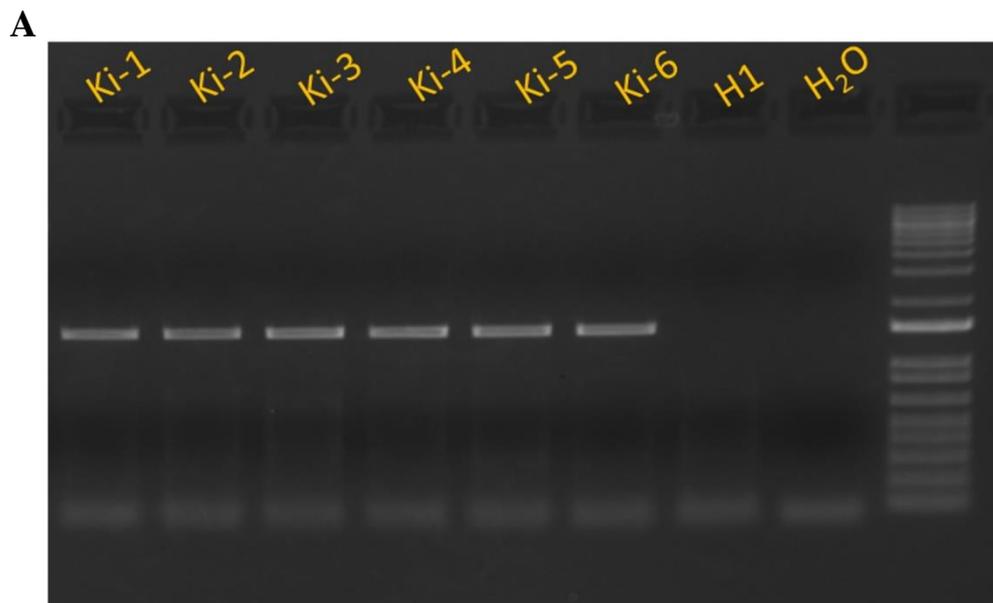
The 20 single-cell clones selected for puromycin resistance, were tested for Cre-ERT2 expression both in real Time PCR and in western blot. The clone N2 was then selected to proceed with the next steps of recombination.



**Fig. 35.** Evaluation of Cre-ERT2 expression in the different clones via Real Time PCR (**A**) and western blot (**B**).

The cells that successfully integrated the Flag-BCOR construct (Fig. 35 B) were selected via Neomycin treatment, and the expression of the recombinant protein in the single-cell clones was checked via western blot.

The selected clones presenting both the chimeric gene Cre-ERT2 and the Flag-tagged BCOR will be subjected to the insertion of the “flippable” construct in the terminal portion of BCOR. Thus, the obtained model will be used to study cell stemness maintenance and differentiation.



**Fig. 36.** Screening of *Flag-BCOR* clones via PCR (A) and western blot (B).

## 5. DISCUSSIONS AND FUTURE PERSPECTIVES

Clear cell sarcoma of the kidney (CCSK) is the second most common childhood renal tumor after the much more frequent Wilms tumor (WT) <sup>[7]</sup>. The distinctive characterization of the tumor is strongly limited by the macroscopic analogies with WT, of which in fact it was considered an aggressive variant until a few decades ago <sup>[38,39,40]</sup>. The identification of the molecular mechanisms underlying the genesis of CCSK therefore represents an important goal for the recognition of unique molecular markers. This need has taken on new strength especially in the light of the identification of a mutation that presumably represents the oncogenic driver in 80-100% of cases, exemplified by the presence of ITD at the level of the last exon of the BCOR gene <sup>[1,52,53]</sup>.

BCOR, acronym for BCL-6 co-repressor, is a gene located in position Xp11.4 coding for different protein isoforms that perform functions dependent on the domains preserved by the splicing process <sup>[64]</sup>. The protein is ubiquitously expressed and interacts with various partners such as transcription factors and other cofactors present within molecular complexes with repressive activity, such as the non-canonical Polycomb Repressive Complex 1.1 (or PRC1.1) . Among the most significant interactions are the ones with the transcription factors BCL-6 and IRF8, the cofactors PCGF1 and KDM2B and the transcriptional activator AF9, which underlines the presence of BCOR in multiple cellular pathways <sup>[66-70]</sup>.

Mutations in BCOR have been found in various pathologies. Cases of malignancies presenting fusion transcripts such as ZC3H7B-BCOR or BCOR-CCNB3 show over-expression of the chimeric protein, similarly to what occurs in CCSK. However, it should be kept in mind that in both fusion genes the PUF1 (PCGF Ub-like Discriminator) domain remains conserved, which instead does not occur in CCSK as the sequence is altered by the presence of the ITD.

Cases of germline loss of function mutations affecting BCOR determine instead, in heterozygosity, the oculo-facio-cardio-dental syndrome (OFCD), inherited in a dominant X-linked manner and lethal in males.

The lack of evidence regarding molecular characterizations of mutations affecting the PUF<sub>D</sub> domain of BCOR in other tumor subtypes has further underlined the need for an in-depth study of CCSK in the context of the altered regulation of pathways involving this protein, and the need to obtain an in vitro model to evaluate its behavior in relation to the other cofactors.

In order to study the complex network of interactions in which BCOR takes part and the functionalities of the mutated protein in a cellular context, two different study approaches have been implemented: a first in-vitro model in HEK-293 cells and a more sophisticated model in human stem cells (H1).

In parallel, the PRC1.1 complex reconstruction in HEK-293 FT cells by overexpressing the core subunits of the complex along with different BCOR variants permitted to gain a first insight on the molecular interactions of BCOR-ITD.

To obtain a first in vitro model for clear cell sarcoma of the kidney, gene editing was attempted by exploiting the characteristics of the CRISPR/Cas9 system for the insertion of a 90 bp ITD at the level of the last exon of BCOR in HEK-293 cells. The system was designed to allow multiple selection steps in order to increase the efficiency.

The selection process, starting from the bulk of transfected cells, led to the isolation of 99 single-cell clones, of which 3 were selected and further characterized. Sanger sequencing revealed the successful insertion of the BCOR-ITD plus P2A-FGP construct in homozygosis in only one clone (C.2F) while C.17 and C.19 clones presented, other than the full insertion of the construct in one allele, a partial insertion of the GFP sequence in the last exon of BCOR. A single-cell clone obtained from cells transfected with a scramble guide vector (SCR1) was used as a control. Also, the previously obtained C.56 clone presented the deletion of the last exon of BCOR, consisting in the loss of a consistent portion of the PUF<sub>D</sub> domain, thus making of it a functional K.O..

The evaluation of GFP expression in the selected clones via flow cytometer analysis revealed comparable expression of the fluorescent marker. Notably, the presence of a double peak in the clone C.19 led to the thought that it could be due to the presence of two populations derived from an erroneous cloning.

Similar range of BCOR expression levels in the different ITD clones and the fact that C.17 clone presented the same recombinational pattern, support the double population hypothesis within clone C.19.. This fact could also explain the presence in the C.19 clone of only the 65% of diploid cells for the X chromosome, as showed in FISH.

The C.56 clone, on the other end, shown an almost complete diploidy for the X chromosome, just like the homozygote clone C.2F. It is also noteworthy that the C.56 clone, missing of the last BCOR exon, also shown an almost complete absence of BCOR expression.

Surprisingly, the ITD clones, as well as the K.O. one, didn't show any growth advantages compared to the controls in normal culture conditions. However, when forced to grow in no-adhesion conditions, the homozygote clone showed a severe growth deficit compared to the control. The heterozygote clones, on the other hand, showed an intermediate phenotype, with a slightly lower growth rate than the wild type, but still presenting the same small spheroids seen in the C.2F.

Contrary to what was seen for the ITD clones, the K.O. clone shown instead a significantly increased growth rate, generating spheroids' macro-aggregates.

The invasion assays permitted to further evaluate the C2F adhesion-dependent growth and compare it to the c56. The significant increased invasion capacity of the homozygote clone compared to both the control and the K.O. clones suggest that BCOR-ITD may be increasing the capacity of the cells to interact with the extra-cellular matrix and to move within it.

The WTS data of the ITD clones were compared with both the wild type controls and the K.O. clone, demonstrating the existence on an actually homogenous expression profile of the BCOR-ITD clones, which is different according to the presence in homozygosis or heterozygosis of the insertion. The K.O. clone similarly showed a peculiar expression profile different from that of the ITD clones.

The enrichment analysis, carried out using both public and cured data-sets, led to the identification of around 1400 differentially expressed transcripts with a p-value lower than 0,05 between the ITD clones and the controls. Amongst these, the targets of the PRC2 complex's methyl-transferase activity stand out in terms of significance, denoting an altered function of the repressor complex in line with the well-known crosstalk with the activity of the PRC1 or PRC1.1 complexes (in which BCOR participates), and further supporting the relevance of the cellular model, that recapitulates CCSK expression profile just by the insertion of the BCOR-ITD.

Other significantly enriched gene sets in the ITD clones involved the pathway downstream of AKT and WNT activation, and also genes involved in cellular adhesion and interaction with the extra-cellular matrix, which lines up with the phenotype shown in the previous experiments.

It's important to note the different behavior of the K.O. clone, especially from the homozygote clone.

Especially, while the homozygote clone in particular showed an high expression of the genes involved in the above mentioned pathways, the K.O. clone had a little to none expression of those genes, underlying the differences between the two.

These data suggest that the presence of internal tandem duplications in the last exon of BCOR could give birth to a *gain-of-function* of the protein instead of leading to a *loss-of-function*, contrary to what was hypothesized in literature.

Furthermore, the great similarity of the expression profiles of the characterized clones and those reported in the literature regarding CCSK, makes this a reliable cellular model for the study of the pathology.

In order to investigate the possibility of a differential molecular signature between metastatic and localized BCOR-ITD positive CCSKs at the diagnosis, the WTS data from 8 CCSK patients (counting 2 metastatic cases) were compared.

The unsupervised analysis showed an evident clustering of samples based on the type of ITD carried, with a clear segregation between the group of three CCSKs characterized by ITD-1 and that of four CCSKs characterized by ITD-3 on the second principal component.

The analysis of the specific molecular signature for metastatic cases highlighted the over-expression of FGF3, VEGFA, SPP1, ADM and JUND, whose functional role may be correlated with the tendency to metastasize. In particular, FGF3 showed an expression 20 times higher in the metastatic samples compared to the localized tumors, increasing the likelihood that it could be casually involved in the acquisition of a more aggressive phenotype.

The functional role of FGF3 was investigated in-vitro by analyzing the induced cellular migration of the before-mentioned homozygote clone, as a model reproducing the oncogenic hit of CCSK. This experiment showed a specific and significant increase in cell migration over the untreated cell lines, with a significant higher induction of cell migration on the BCOR-ITD with respect to the SCR clone.

Therefore, FGF3 proved to be involved in the gaining of an aggressive phenotype, although is unclear whether its increased level of transcripts found in samples of metastatic tumors is directly related to either an autocrine or a paracrine signaling loops.

Moreover, it was shown that the invasive phenotype induction generated by the overexpression of FGF3 is independent from the type of ITD present, leaving room for speculation about the correlation between FGF3 signaling pathway and the altered BCOR-ITD's activity. This also proposes FGF3 as a marker of interest to monitor CCSK aggressiveness.

PRC1.1 complex reconstruction by transient expression of the core factors showed that different ITDs have different capacities of retaining the complex's core factor, contrary to the  $\Delta$ PUFD variant which, missing the binding domain for PCGF1, couldn't carry out the indispensable function of molecular bridge between PCGF1 and the ubiquitin-ligase activity protein RING1B.

This result strongly supports the hypothesis that the presence of the internal tandem duplications in the last region of BCOR gene cause structural changes in the PUFD domain, and thus are responsible of an altered protein function. Moreover, the comparison with the BCOR  $\Delta$ PUFD variant revealed that the presence of the ITDs cause only a partial loss of BCOR's binding capacity, thus giving less credit to the *loss-of-function* hypothesis.

Also, it is interesting to note that every BCOR-ITD variant showed a peculiar retention pattern, different from the others. These data support the fact that this kind of duplications could vary BCOR binding patterns in a different way, leading to different BCOR functions and thus a peculiar expression profile (as seen in Fiore, Taddia et al.).

Future analysis in mass spectrometry will reveal the whole binding pattern of the different BCOR-ITD variants, providing more evidences that will allow a better understanding of its molecular functions.

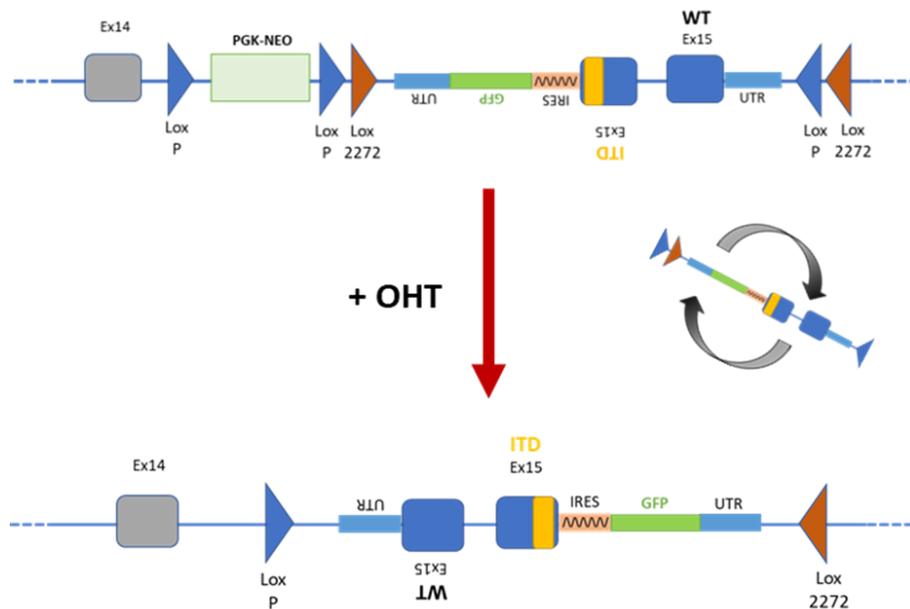
In order to study BCOR-ITD functions in cell stemness and differentiation, it has proven to be necessary to generate a different cellular model from the previous one.

The HEK-293 cell line was indeed not suitable for this purpose, being a line of immortalized cells with an unstable aneuploidy that vary across the population.

The elected cell line for the new model was the human stem cell line H1, in which it was designed the generation of an inducible model for the expression of BCOR-ITD.

Unfortunately, due to the required multi-step insertion of different construct in the selected cells it is still ongoing. The first step required the insertion of the Cre-ERT2 chimeric gene in the AAVS1 locus

under the endogenous promoter. Due to the estrogen receptor portion, the generated protein will be localized on the cellular surface and will be transferred into the nucleus only in response to Tamoxifen (OHT). Here, the Cre enzymatic portion will recognize the Lox sequences and generate the inversion of the construct localized in the terminal portion of the BCOR gene (Fig.34).



**Fig. 34.** Scheme of inducible BCOR-ITD before and after addition of hydroxy tamoxifen (OHT).

Once ultimated this model will permit the study of BCOR-ITD functions not only in the pluripotency maintenance but also in mesenchymal differentiation and finally in the generation of kidney organoids, in order to find a possible correlation between the molecular functions of the mutated BCOR and the CCSK phenotype.

In short, the work carried out in this thesis has set itself the objective of investigating the molecular mechanisms underlying the genesis of clear cell sarcomas of the kidney, approaching the study with different methodologies. A first cellular model in HEK-293 allowed to obtain important information about the functionality of BCOR, suggesting that the presence of internal tandem duplications generates an altered activity of the transcriptional repressor which is very different from a loss-of-

function, as hypothesized so far in the literature. It has also been observed that the function of BCOR within the PRC1 complex varies with the variation of the ITDs present within the terminal part of the gene. Moreover, it proved that BCOR-ITD is able to induce an efficient interaction with the ECM components, and to enhance invasive capacities.

The future direction will be towards the study of the functionality of BCOR-ITD in a context more similar to that of origin of the CCSKs, evaluating, thanks to the inducible model being obtained, both the specific interactome and the phenotypic consequences caused by the mutation.

## 6. BIBLIOGRAPHY

1. Ueno-Yokohata H, Okita H, Nakasato K, et al. Consistent in-frame internal tandem duplications of BCOR characterize clear cell sarcoma of the kidney. *Nat Genet* 2015;47(8):861-863.
2. Steliarova-Foucher E, Stiller C, Kaatsch P, et al. Trends in childhood cancer incidence in Europe, 1970-99. *Lancet* 2005;365(9477):2088.
3. Pritchard-Jones K, J. D. *Renal Tumors of Childhood: Biology and Therapy*. Berlin: Springer-Verlag; 2014.
4. Lowe, L.H., et al., *Pediatric renal masses: Wilms tumor and beyond*. *Radiographics*, 2000. **20**(6): p. 1585-603.
5. Vujanić, G., *Renal tumours in early life*. *Current diagnostic pathology*, 2006. **12**(3): p. 210-219.
6. Charles AK, Vujanić GM, Berry PJ. Renal tumours of childhood. *Histopathology* 1998;32(4):293-309.
7. JN E, G S, JI E, et al. *World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs* 2004.
8. Vujanić GM, Gessler M, Ooms AHAG, Collini P, Coulomb-l'Hermine A, D'Hooghe E, de Krijger RR, Perotti D, Pritchard-Jones K, Vokuhl C, van den Heuvel-Eibrink MM, Graf N; International Society of Paediatric Oncology–Renal Tumour Study Group (SIOP–RTSG). The UMBRELLA SIOP-RTSG 2016 Wilms tumour pathology and molecular biology protocol. *Nat Rev Urol*. 2018 Nov;15(11):693-701. doi: 10.1038/s41585-018-0100-3. Erratum in: *Nat Rev Urol*. 2019 Sep;16(9):563.
9. Davidoff AM. Wilms' tumor. *Curr Opin Pediatr* 2009;21(3):357-364.
10. Irtan S, Ehrlich PF, Pritchard-Jones K. Wilms tumor: "State-of-the-art" update, 2016. *Semin Pediatr Surg* 2016;25(5):250-256.
11. Kieran K, Ehrlich PF. Current surgical standards of care in Wilms tumor. *Urol Oncol* 2016;34(1):13-23.
12. Dome JS, Perlman EJ, Graf N. Risk stratification for wilms tumor: current approach and future directions. *Am Soc Clin Oncol Educ Book* 2014:215-223.
13. Dome JS, Fernandez CV, Mullen EA, et al. Children's Oncology Group's 2013 blueprint for research: renal tumors. *Pediatr Blood Cancer* 2013;60(6):994-1000.
14. Breslow N, Olshan A, Beckwith JB, et al. Epidemiology of Wilms tumor. *Med Pediatr Oncol* 1993;21(3):172-181.
15. Segers H, van den Heuvel-Eibrink MM, Pritchard-Jones K, et al. Management of adults with Wilms' tumor: recommendations based on international consensus. *Expert Rev Anticancer Ther* 2011;11(7):1105-1113.
16. Indolfi P, Jenkner A, Terenziani M, et al. Synchronous bilateral Wilms tumor: a report from the Associazione Italiana Ematologia Oncologia Pediatrica (AIEOP). *Cancer* 2013;119(8):1586-1592.
17. R S, M H, PS S, et al. *Extrarenal Wilms' Tumor: Challenges in Diagnosis, Embryology, Treatment and Prognosis*. Brisbane (AU): Codon; 2016.
18. Miser JS, Tournade MF. The management of relapsed Wilms tumor. *Hematol Oncol Clin North Am* 1995;9(6):1287-1302.

19. Rivera, M.N. and D.A. Haber, *Wilms' tumour: connecting tumorigenesis and organ development in the kidney*. Nat Rev Cancer, 2005. **5**(9): p. 699-712.
20. Breslow, N., et al., *Epidemiology of Wilms tumor*. Med Pediatr Oncol, 1993. **21**(3): p. 172-81.
21. Beckwith, J.B., N.B. Kiviat, and J.F. Bonadio, *Nephrogenic rests, nephroblastomatosis, and the pathogenesis of Wilms' tumor*. Pediatr Pathol, 1990. **10**(1-2): p. 1-36.
22. Popov, S.D., N.J. Sebire, and G.M. Vujanic, *Wilms' Tumour - Histology and Differential Diagnosis*, in *Wilms Tumor*, M.M. van den Heuvel-Eibrink, Editor. 2016, Codon Publications Copyright: The Authors.: Brisbane (AU).
23. Hou, L.T. and R.L. Holman, *Bilateral nephroblastomatosis in a premature infant*. J Pathol Bacteriol, 1961. **82**: p. 249-55.
24. Kacar A, Azili MN, Cihan BS, et al. Metanephric stromal tumor: a challenging diagnostic entity in children. J Pediatr Surg 2011;46(12):e7-e10.
25. Sebire NJ, Vujanic GM. Paediatric renal tumours: recent developments, new entities and pathological features. Histopathology 2009;54(5):516-528.
26. Beckwith JB, Zuppan CE, Browning NG, et al. Histological analysis of aggressiveness and responsiveness in Wilms' tumor. Med Pediatr Oncol 1996;27(5):422-428.
27. Vujančić GM, Sandstedt B, Harms D, et al. Revised International Society of Paediatric Oncology (SIOP) working classification of renal tumors of childhood. Med Pediatr Oncol 2002;38(2):79-82.
28. Dome JS, Cotton CA, Perlman EJ, et al. Treatment of anaplastic histology Wilms' tumor: results from the fifth National Wilms' Tumor Study. J Clin Oncol 2006;24(15):2352-2358.
29. Scott RH, Stiller CA, Walker L, et al. Syndromes and constitutional chromosomal abnormalities associated with Wilms tumour. J Med Genet 2006;43(9):705-715.
30. Rump P, Zeegers MP, van Essen AJ. Tumor risk in Beckwith-Wiedemann syndrome: A review and meta-analysis. Am J Med Genet A 2005;136(1):95-104.
31. Dome JS, Coppes MJ. Recent advances in Wilms tumor genetics. Curr Opin Pediatr 2002;14(1):5-11.
32. Brok J, Treger TD, Gooskens SL, et al. Biology and treatment of renal tumours in childhood. Eur J Cancer 2016;68:179-195.
33. Ruteshouser EC, Robinson SM, Huff V. Wilms tumor genetics: mutations in WT1, WTX, and CTNNB1 account for only about one-third of tumors. Genes Chromosomes Cancer 2008;47(6):461-470.
34. Torrezan GT, Ferreira EN, Nakahata AM, et al. Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. Nat Commun 2014;5:4039.
35. Gratias EJ, Dome JS, Jennings LJ, et al. Association of Chromosome 1q Gain With Inferior Survival in Favorable-Histology Wilms Tumor: A Report From the Children's Oncology Group. J Clin Oncol 2016;34(26):3189-3194.
36. Grundy PE, Breslow NE, Li S, et al. Loss of heterozygosity for chromosomes 1p and 16q is an adverse prognostic factor in favorable-histology Wilms tumor: a report from the National Wilms Tumor Study Group. J Clin Oncol 2005;23(29):7312-7321.
37. Gooskens SL, Furtwängler R, Vujanic GM, et al. Clear cell sarcoma of the kidney: a review. Eur J Cancer 2012;48(14):2219-2226.
38. Argani P, Perlman EJ, Breslow NE, et al. Clear cell sarcoma of the kidney: a review of 351 cases from the National Wilms Tumor Study Group Pathology Center. Am J Surg Pathol 2000;24(1):4-18.

39. Saula PW, Hadley GP. Pediatric non-Wilms' renal tumors: a third world experience. *World J Surg* 2012;36(3):565-572.
40. Manchanda V, Mohta A, Khurana N, et al. Bilateral clear cell sarcoma of the kidney. *J Pediatr Surg* 2010;45(9):1927-1930.
41. Zekri W, Yehia D, Elshafie MM, et al. Bilateral clear cell sarcoma of the kidney. *J Egypt Natl Canc Inst* 2015;27(2):97-100.
42. Wong MK, Ng CCY, Kuick CH, et al. Clear cell sarcomas of the kidney are characterised by BCOR gene abnormalities, including exon 15 internal tandem duplications and BCOR-CCNB3 gene fusion. *Histopathology* 2018;72(2):320-329.
43. Gooskens SL, Furtwängler R, Spreafico F, et al. Treatment and outcome of patients with relapsed clear cell sarcoma of the kidney: a combined SIOP and AIEOP study. *Br J Cancer* 2014;111(2):227-233.
44. Boo YJ, Fisher JC, Haley MJ, et al. Vascular characterization of clear cell sarcoma of the kidney in a child: a case report and review. *J Pediatr Surg* 2009;44(10):2031-2036.
45. WM M, JB B, GM F. Tumors of the kidney. Washington D.C.:Armed Forces Institute of Pathology 1994.
46. Balarezo FS, Joshi VV. Clear cell sarcoma of the pediatric kidney: detailed description and analysis of variant histologic patterns of a tumor with many faces. *Adv Anat Pathol* 2001;8(2):98-108.
47. Haas JE, Bonadio JF, Beckwith JB. Clear cell sarcoma of the kidney with emphasis on ultrastructural studies. *Cancer* 1984;54(12):2978-2987.
48. Mirkovic J, Calicchio M, Fletcher CD, et al. Diffuse and strong cyclin D1 immunoreactivity in clear cell sarcoma of the kidney. *Histopathology* 2015;67(3):306-312.
49. Ueno H, Okita H, Akimoto S, et al. DNA methylation profile distinguishes clear cell sarcoma of the kidney from other pediatric renal tumors. *PLoS One* 2013;8(4):e62233.
50. Kao YC, Sung YS, Zhang L, et al. Recurrent BCOR Internal Tandem Duplication and YWHAE-NUTM2B Fusions in Soft Tissue Undifferentiated Round Cell Sarcoma of Infancy: Overlapping Genetic Features With Clear Cell Sarcoma of Kidney. *Am J Surg Pathol* 2016;40(8):1009-1020.
51. Kenny C, Bausenwein S, Lazaro A, et al. Mutually exclusive BCOR internal tandem duplications and YWHAE-NUTM2 fusions in clear cell sarcoma of kidney: not the full story. *J Pathol* 2016;238(5):617-620.
52. Astolfi A, Melchionda F, Perotti D, et al. Whole transcriptome sequencing identifies BCOR internal tandem duplication as a common feature of clear cell sarcoma of the kidney. *Oncotarget* 2015;6(38):40934-40939.
53. Roy A, Kumar V, Zorman B, et al. Recurrent internal tandem duplications of BCOR in clear cell sarcoma of the kidney. *Nat Commun* 2015;6:8891.
54. Karlsson J, Valind A, Gisselsson D. BCOR internal tandem duplication and YWHAE-NUTM2B/E fusion are mutually exclusive events in clear cell sarcoma of the kidney. *Genes Chromosomes Cancer* 2016;55(2):120-123.
55. Karlsson, J., A. Valind, and D. Gisselsson, *BCOR internal tandem duplication and YWHAE-NUTM2B/E fusion are mutually exclusive events in clear cell sarcoma of the kidney*. *Genes Chromosomes Cancer*, 2016. **55**(2): p. 120-3.
56. Karlsson, J., et al., Aberrant epigenetic regulation in clear cell sarcoma of the kidney featuring distinct DNA hypermethylation and EZH2 overexpression. *Oncotarget*, 2016. **7**(10): p. 11127-36.

57. Cutcliffe, C., et al., *Clear cell sarcoma of the kidney: up-regulation of neural markers with activation of the sonic hedgehog and Akt pathways*. Clin Cancer Res, 2005. **11**(22): p. 7986-94.
58. Karlsson, J., et al., *Clear cell sarcoma of the kidney demonstrates an embryonic signature indicative of a primitive nephrogenic origin*. Genes Chromosomes Cancer, 2014. **53**(5): p. 381-91. 178
59. Gooskens, S.L., et al., *TCF21 hypermethylation in genetically quiescent clear cell sarcoma of the kidney*. Oncotarget, 2015. **6**(18): p. 15828-41.
60. Huang, C.C., et al., *Classification of malignant pediatric renal tumors by gene expression*. Pediatr Blood Cancer, 2006. **46**(7): p. 728-38.
61. Wang, Z., et al., *Revealing the role of VEGFA in clear cell sarcoma of the kidney by protein-protein interaction network and significant pathway analysis*. Oncol Lett, 2016. **11**(2): p. 953-958.
62. Volkel, P., et al., *Diverse involvement of EZH2 in cancer epigenetics*. Am J Transl Res, 2015. **7**(2): p. 175-93.
63. Poirier, J.T., et al., *DNA methylation in small cell lung cancer defines distinct disease subtypes and correlates with high expression of EZH2*. Oncogene, 2015. **34**(48): p. 5869-78
64. Arab, K., et al., *Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A*. Mol Cell, 2014. **55**(4): p. 604-14.
65. Huynh KD, Fischle W, Verdin E, et al. BCoR, a novel corepressor involved in BCL-6 repression. Genes Dev 2000;14(14):1810-1823.
66. Pagan JK, Arnold J, Hanchard KJ, et al. A novel corepressor, BCoR-L1, represses transcription through an interaction with CtBP. J Biol Chem 2007;282(20):15248-15257.
67. Tiacci E, Grossmann V, Martelli MP, et al. The corepressors BCOR and BCORL1: two novel players in acute myeloid leukemia. Haematologica 2012;97(1):3-5.
68. Wong SJ, Gearhart MD, Taylor AB, et al. KDM2B Recruitment of the Polycomb Group Complex, PRC1.1, Requires Cooperation between PCGF1 and BCORL1. Structure 2016;24(10):1795-1801.
69. Junco SE, Wang R, Gaipa JC, et al. Structure of the polycomb group protein PCGF1 in complex with BCOR reveals basis for binding selectivity of PCGF homologs. Structure 2013;21(4):665-671.
70. Srinivasan RS, de Erkenez AC, Hemenway CS. The mixed lineage leukemia fusion partner AF9 binds specific isoforms of the BCL-6 corepressor. Oncogene 2003;22(22):3395-3406.
71. Wang, Z., et al., *A Non-canonical BCOR-PRC1.1 Complex Represses Differentiation Programs in Human ESCs*. Cell Stem Cell, 2018. 22.
72. Cardenas MG, Oswald E, Yu W, et al. The Expanding Role of the BCL6 Oncoprotein as a Cancer Therapeutic Target. Clin Cancer Res 2017;23(4):885-893.
73. Miles RR, Crockett DK, Lim MS, et al. Analysis of BCL6-interacting proteins by tandem mass spectrometry. Mol Cell Proteomics 2005;4(12):1898-1909.
74. Wamstad JA, Corcoran CM, Keating AM, et al. Role of the transcriptional corepressor Bcor in embryonic stem cell differentiation and early embryonic development. PLoS One 2008;3(7):e2814.
75. Cao Q, Gearhart MD, Gery S, et al. BCOR regulates myeloid cell proliferation and differentiation. Leukemia 2016;30(5):1155-1165.
76. Sturm D, Orr BA, Toprak UH, et al. New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. Cell 2016;164(5):1060-1072.

77. Panagopoulos I, Thorsen J, Gorunova L, et al. Fusion of the ZC3H7B and BCOR genes in endometrial stromal sarcomas carrying an X;22-translocation. *Genes Chromosomes Cancer* 2013;52(7):610-618.
78. Micci F, Gorunova L, Agostini A, et al. Cytogenetic and molecular profile of endometrial stromal sarcoma. *Genes Chromosomes Cancer* 2016;55(11):834-846.
79. García-Sanz P, Triviño JC, Mota A, et al. Chromatin remodelling and DNA repair genes are frequently mutated in endometrioid endometrial carcinoma. *Int J Cancer* 2017;140(7):1551-1563.
80. Peters TL, Kumar V, Polikepahad S, et al. BCOR-CCNB3 fusions are frequent in undifferentiated sarcomas of male children. *Mod Pathol* 2015;28(4):575-586.
81. Machado I, Navarro L, Pellin A, et al. Defining Ewing and Ewing-like small round cell tumors (SRCT): The need for molecular techniques in their categorization and differential diagnosis. A study of 200 cases. *Ann Diagn Pathol* 2016;22:25-32.
82. Santiago T, Clay MR, Allen SJ, et al. Recurrent BCOR internal tandem duplication and BCOR or BCL6 expression distinguish primitive myxoid mesenchymal tumor of infancy from congenital infantile fibrosarcoma. *Mod Pathol* 2017;30(6):884-891.
83. Shern JF, Chen L, Chmielecki J, et al. Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. *Cancer Discov* 2014;4(2):216-231.
84. Seki M, Nishimura R, Yoshida K, et al. Integrated genetic and epigenetic analysis defines novel molecular subgroups in rhabdomyosarcoma. *Nat Commun* 2015;6:7557.
85. Temming P, Corson TW, Lohmann DR. Retinoblastoma tumorigenesis: genetic and epigenetic changes walk hand in hand. *Future Oncol* 2012;8(5):525-528.
86. Pugh TJ, Weeraratne SD, Archer TC, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* 2012;488(7409):106-110.
87. Damm F, Chesnais V, Nagata Y, et al. BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders. *Blood* 2013;122(18):3169-3177.
88. Totoki Y, Tatsuno K, Yamamoto S, et al. High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* 2011;43(5):464-469.
89. Golbabapour, S., et al., Gene silencing and Polycomb group proteins: an overview of their structure, mechanisms and phylogenetics. *Omics : a journal of integrative biology*, 2013. 17(6): p. 283-296.
90. Kassis, J.A., J.A. Kennison, and J.W. Tamkun, Polycomb and trithorax group genes in *Drosophila*. *Genetics*, 2017. 206(4): p. 1699-1725.
91. Ringrose, L. and R. Paro, Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.*, 2004. 38: p. 413-443.
92. Rickels, R., et al., An evolutionary conserved epigenetic mark of polycomb response elements implemented by Trx/MLL/COMPASS. *Molecular cell*, 2016. 63(2): p. 318-328.
93. Schuettengruber, B., et al., Genome regulation by polycomb and trithorax proteins. *Cell*, 2007. 128(4): p. 735-745.
94. Boyer, L.A., et al., Polycomb complexes repress developmental regulators in murine embryonic stem cells. *nature*, 2006. 441(7091): p. 349-353.
95. Levine, S.S., et al., The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Molecular and cellular biology*, 2002. 22(17): p. 6070-6078.
96. Gao Z, Zhang J, Bonasio R, et al. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* 2012;45(3):344-356.

97. Morey L, Aloia L, Cozzuto L, et al. RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. *Cell Rep* 2013;3(1):60-69.
98. Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* 157, 1445–1459.
99. Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* 45, 344–356.
100. Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873–878.
101. Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* 1, e00205.
102. Huang, Y., Zhao, W., Wang, C., Zhu, Y., Liu, M., Tong, H., Xia, Y., Jiang, Q., and Qin, J. (2018). Combinatorial Control of Recruitment of a Variant PRC1.6 Complex in Embryonic Stem Cells. *Cell Rep.* 22, 3032–3043.
103. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet.* 14, e1007193.
104. Scelfo, A., Fernandez-Perez, D., Tamburri, S., Zanotti, M., Lavarone, E., Soldi, M., Bonaldi, T., Ferrari, K.J., and Pasini, D. (2019). Functional Landscape of PCGF Proteins Reveals Both RING1A/B-Dependent-and RING1A/BIndependent-Specific Activities. *Mol. Cell* 74, 1037–1052.
105. Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* 53, 49–62.
106. Lavarone, E., Barbieri, C.M., and Pasini, D. (2019). Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. *Nat. Commun.* 10, 1679.
107. Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* 32, 503–518.
108. Shen, X., Liu, Y., Hsu, Y.J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* 32, 491–502.
109. Beringer, M., Pisano, P., Di Carlo, V., Blanco, E., Chammas, P., Vizaín, P., Gutiérrez, A., Aranda, S., Payer, B., Wierer, M., and Di Croce, L. (2016). EPOP Functionally Links Elongin and Polycomb in Pluripotent Stem Cells. *Mol. Cell* 64, 645–658.
110. Conway, E., Jerman, E., Healy, E., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D.J., Mucha, M., et al. (2018). A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. *Mol. Cell* 70, 408–421.
111. Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., et al. (2016). Jarid2 binds mono-

- ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat. Commun.* 7, 13661.
112. Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W., Müller, C.W., Vermeulen, M., and Muller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* 21, 569–571.
  113. Blackledge, N.P., Rose, N.R., and Klose, R.J. (2015). Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat. Rev. Mol. Cell Biol.* 16, 643–649.
  114. Tamburri S, Lavarone E, Fernández-Pérez D, Conway E, Zanotti M, Manganaro D, Pasini D. Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Mol Cell.* 2020 Feb 20;77(4):840-856.e5.
  115. Zhang HX, Zhang Y, Yin H. Genome Editing with mRNA Encoding ZFN, TALEN, and Cas9. *Mol Ther.* 2019 Apr 10;27(4):735-746. doi: 10.1016/j.ymthe.2019.01.014. Epub 2019 Jan 25. PMID: 30803822; PMCID: PMC6453514.
  116. Tebas P, Stein D, Tang WW, et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med* 2014;370(10):901-910.
  117. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014;157(6):1262-1278.
  118. Urnov, F.D., et al., Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, 2010. 11(9): p. 636-646.
  119. Zhang, M., et al., TALE: a tale of genome editing. *Progress in Biophysics and Molecular Biology*, 2014. 114: p. 25-32.
  120. Mojica FJ, Díez-Villaseñor C, García-Martínez J, et al. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 2005;60(2):174-182.
  121. Jansen R, Embden JD, Gaastra W, et al. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 2002;43(6):1565-1575.
  122. Haft DH, Selengut J, Mongodin EF, et al. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 2005;1(6):e60.
  123. Makarova KS, Haft DH, Barrangou R, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 2011;9(6):467-477.
  124. Brouns SJ, Jore MM, Lundgren M, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 2008;321(5891):960-964.
  125. Hale CR, Zhao P, Olson S, et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 2009;139(5):945-956.
  126. Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337(6096):816-821.
  127. Gasiunas G, Barrangou R, Horvath P, et al. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 2012;109(39):E2579-2586.
  128. Ran FA, Hsu PD, Wright J, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 2013;8(11):2281-2308.
  129. Wyman C, Kanaar R. DNA double-strand break repair: all's well that ends well. *Annu Rev Genet* 2006;40:363-383.
  130. Prakash V, Moore M, Yáñez-Muñoz RJ. Current Progress in Therapeutic Gene Editing for Monogenic Diseases. *Mol Ther* 2016;24(3):465-474.

131. Cox DB, Platt RJ, Zhang F. Therapeutic genome editing: prospects and challenges. *Nat Med* 2015;21(2):121-131.
132. Astolfi A, Fiore M, Melchionda F, Indio V, Bertuccio SN, Pession A. BCOR involvement in cancer. *Epigenomics*. 2019 May;11(7):835-855. doi: 10.2217/epi-2018-0195. Epub 2019 May 31. PMID: 31150281; PMCID: PMC6595546.
133. White MK, Khalili K. CRISPR/Cas9 and cancer targets: future possibilities and present challenges. *Oncotarget* 2016;7(11):12305-12317.
134. Fiore M, Taddia A, Indio V, Bertuccio SN, Messelodi D, Serravalle S, Bandini J, Spreafico F, Perotti D, Collini P, Di Cataldo A, Pasquinelli G, Chiarini F, Fois M, Melchionda F, Pession A, Astolfi A. Molecular Signature of Biological Aggressiveness in Clear Cell Sarcoma of the Kidney (CCSK). *Int J Mol Sci*. 2023 Feb 13;24(4):3743. doi: 10.3390/ijms24043743. PMID: 36835166; PMCID: PMC9964999.
135. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25(9):1105-1111.
136. KEGG - Kyoto Encyclopedia of Genes and Genomes. [cited 2017/01/05]; Available from: <http://www.genome.jp/kegg/>.
137. <https://ocg.cancer.gov/programs/target>
138. WEB-based GEne SeT AnaLysis Toolkit. [cited 2017 01/05]; Available from: <http://www.webgestalt.org>.