

# Genomic instability and genetic heterogeneity in neuroblastoma

Niloufar Javanmardi

Department of Pathology and Clinical  
Genetics, Institute of Biomedicine at  
Sahlgrenska Academy  
University of Gothenburg

Gothenburg, Sweden 2017



UNIVERSITY OF  
GOTHENBURG

Cover illustration by Fatemeh Arabi

*Genomic instability and genetic heterogeneity in neuroblastoma - From improving risk stratification to identifying actionable genomic alterations*

© 2017 Niloufar Javanmardi  
niloufar.javanmardi@gu.se

ISBN 978-91-629-0336-7

Printed by Ineko  
Gothenburg, Sweden 2017

به نام خدای دشت نیلوفر چشم به راه آشنا

*To my wonderful family*

On ne pense que par image. Si tu veux être  
philosophe écris des romans.  
(Albert Camus)

# Dissertation abstract

Javanmardi, N., 2017, Genomic instability and genetic heterogeneity in neuroblastoma  
Department of Pathology and Clinical Genetics, Institute of Biomedicine at  
Sahlgrenska Academy, University of Gothenburg, Sweden

Neuroblastoma (NB), a tumour of the sympathetic nervous system and the most common malignant disease of early childhood, is responsible for 9% of paediatric cancer related deaths. Aggressive NB still constitutes a major clinical problem with survival rates of about 35%. It is therefore of great clinical interest to further study the biological parameters that can (i) better classify tumours so that the children may be given the right treatment (ii) identify new actionable targets.

**Aim** - the objective of this thesis was to explore genes and chromosomal regions with potential involvement in the initiation/progression of NB that can be used for improved patient stratification.

**Results** – In paper I and III we detected point mutations in the tyrosine kinase domain of the *ALK* oncogene. Minor population of cells with *ALK* mutations were detected with massive parallel deep DNA sequencing. It is likely that early detection of subclones with *ALK* mutation is critical in treatment of these tumours with recently derived small molecule ALK inhibitors. We propose increased serial sampling of tumour material from high-risk NB tumours and analysis with the new sequencing techniques.

In paper II we observed that the distal part of chromosome arm 2p often is subjected to gain of an extra copy – i.e. 2p-gain. Interestingly, this region contains three genes, *ALKAL2*, *MYCN* and *ALK*, of strong importance for NB development. We suggest that the gain of this “cassette” of genes is beneficial to the NB tumor pathogenesis with potential to aid in therapeutical intervention.

In the last study, paper IV, we analysed the high-risk 11q-deleted NB tumours. We show that 11q-deleted tumours with and without *MYCN* amplification present different 11q-deletion breakpoint patterns. The detailed analysis of these patterns enabled us to detect genes and chromosomal regions on 11q that may contain tumour suppressors in this severe child cancer subgroup. Furthermore, we propose *DLG2* as a highly interesting 11q candidate NB gene.

**Conclusion** - Our observation of a significant spatiotemporal variation of *ALK* mutations is of utmost importance in clinical practice. *DLG2* stands out as a strong tumor suppressor candidate for the 11q-deleted NBs. It is important to note that the experiments we propose are expected to contribute to precision medicine.

**Keywords** - tumour, neural crest, neuroblastoma, subclone, mutation, relapse, deep sequencing, microarray, 2p, *MYCK*, *ALK*, *ALKAL2*, *11q*, *DLG2*, *CCND1*

# Sammanfattning på svenska

Javanmardi, N, 201, Genomisk instabilitet och genetisk heterogenitet i neuroblastom  
Avdelningen för Patologi och Genetik, Institutionen för Biomedicin, Sahlgrenska  
akademien vid Göteborgs Universitet, Sverige

Neuroblastom (NB) är en barncancerform där tumörer uppkommer i det sympatiska nervsystemet. NB är en av de vanligaste tumörformerna bland unga barn och orsakar 9 % av alla cancerrelaterade dödsfall hos barn. Den mest aggressiva formen av NB är mycket svårbehandlad och överlevnaden för denna grupp endast är cirka 35 %. Därför är det av stor betydelse att studera biologiska faktorer som kan (i) användas för att förbättra klassificeringen av tumörer så att barnen får rätt behandling (ii) identifiera nya möjliga behandlingsmål.

**Syfte** – Syftet med denna avhandling är att utforska förändringar i gener och kromosomala regioner som kan vara involverade i uppkomst/progression av NB och som kan användas till förbättrad patientgruppering inför behandling.

**Resultat** – I arbete I och III studerade vi förekomsten av aktiverande punktmutationer i kinasdomänen hos onkogenen *ALK*. Genom att använda ultradep DNA-sekvensering kunde vi påvisa närvaro av förändringar i *ALK* som bara förekom i en liten andel av cellerna i tumören. Då ett flertal läkemedel som inhiberar *ALK* är under utveckling så tror vi att möjligheten att detektion av subklonala *ALK*-mutationer är av stor betydelse för NB-patienter. För att kunna fånga upp mutationerna i så tidigt skede som möjligt så föreslår vi en mer frekvent provtagning av högrisk-NB och analys med den nya DNA-sekvenseringsteknologin.

I arbete II observerar vi att den distala delen av det kromosomala området 2p ofta förekommer i en extra kopia, ett så kallat 2p-gain. Intressant nog så ligger tre gener; *ALKAL2*, *MYCN* och *ALK*, som alla har stor betydelse för utveckling av NB inom denna region. Vi tror att extrakopior av dessa tre gener tillsammans har en drivande funktion vid uppkomsten av NB och därför kan vara viktigt att ta i beaktande vid val av behandlingsstrategi.

I arbete IV studerade vi gruppen av hög-risk NB med 11q-deletion. Vi visar att tumörer med 11q-deletion har olika mönster av brottspunkter på 11q beroende av om tumörerna också har amplifiering av *MYCN*. Detaljerad analys av brottspunkter och deletioner gör det möjligt att identifiera genomiska områden inom 11q som kan innehålla tumörsuppressorgener i denna aggressiva grupp av NB-tumörer. En mycket intressant kandidat-tumörsuppressorgen som identifierats inom det frekvent deleterade området på 11q är *DLG2*.

**Slutsats** – I våra studier observerar vi skillnader i förekomst av *ALK*-mutationer, både inom tumören och över tid, vilket kan vara av yttersta vikt i den kliniska utvärderingen. Det påvisar vikten av att använda ultrakänsliga metoder som t.ex. den nya DNA-sekvenseringsteknologin men också behovet av ökad provtagningsfrekvens. Vi visar också att *DLG2* utmärker sig som en stark tumörsuppressorkandidat för gruppen av NB-tumörer med 11q-deletion. Det är också viktigt att notera att de analyser som vi föreslår är ett led till individanpassad behandling, dvs att ge rätt behandling till rätt patient.

# List of papers

This thesis is based on the following studies, referred to in the text by their Roman numerals.

- I. Schleiermacher G, **Javanmardi N**, Bernard V, Leroy Q, Cappo J, Rio Frio T, Pierron G, Lapouble E, Combaret V, Speleman F, de Wild B, Djos A, Ora I, Hedborg F, Träger C, Holmqvist BM, Abrahamsson J, Peuchmaur M, Michon J, Janoueix-Lerosey I, Kogner P, Delattre O, Martinsson T; *Emergence of new ALK mutations at relapse of neuroblastoma*. J Clin Oncol. 2014 Sep 1;32(25):2727-34
- II. **Javanmardi N**, Fransson S, Djos A, Umapathy G, Östensson M, Milosevic J, Kogner P, Hallberg B, Martinsson T, and Palmer RH; *The 2p cassette gain in neuroblastoma tumours, A potent combination of ALK, MYCN and the ALK ligand ALKAL2 (FAM150B/AUGa)*. 2017, Manuscript
- III. **Javanmardi N**, Fransson S, Djos A, Sjöberg RM, Lorentzen E, Truvé K, Kogner P, Martinsson T; *Low frequency ALK hotspots mutations in Neuroblastoma tumours detected by ultra deep sequencing: Implications for ALK inhibitor treatment*. 2017, Manuscript
- IV. **Javanmardi N**, Fransson S, Djos A, Sjöberg RM, Östensson M, Bergerall A, Carén H, Beiske K, Palmer RH, Hallberg B, Noguera R, Kogner P and Martinsson T; *Chromosome 11 in high-risk neuroblastoma tumours; A loss and gain pattern indicative of both tumour suppressor and oncogene activity*. 2017, Manuscript

# Content

<b>ABBREVIATIONS .....</b>	<b>9</b>
<b>INTRODUCTION .....</b>	<b>11</b>
<i>BASIC GENETICS</i> .....	11
DNA and genes .....	11
The central dogma of molecular biology .....	12
Genetic variations .....	13
Organization of the genetic material.....	13
<i>EPIGENETICS</i> .....	14
DNA methylation .....	14
Histone Modifications .....	15
RNA-associated silencing .....	15
TAD regulation in the 3D genome .....	15
<i>CANCER</i> .....	16
Oncogenes and tumour suppressor genes .....	16
The two-hit hypothesis .....	16
Epigenetic and cancer .....	17
TAD and cancer .....	17
<i>NEUROBLASTOMA</i> .....	18
Epidemiology .....	18
Symptoms and treatment .....	18
Hereditary neuroblastoma .....	18
Prognostic factors .....	19
Expression of neurotrophin receptors .....	19
Ploidy .....	20
Tumour histology .....	20
Risk stratification.....	20
Chromosomal abnormalities .....	22
1p deletion.....	22
Chromosome arm 2p: MYCN, ALK, ALKAL2.....	23
11q loss.....	24
17q gain.....	24
Other chromosomal aberrations .....	25
<i>ALK</i> (anaplastic lymphoma kinase) .....	25
ALK, MYCN and NB.....	25
ALK as a therapeutic target.....	27

Tyrosine kinase inhibitors .....	27
<b>OBJECTIVES.....</b>	<b>28</b>
Paper I.....	28
Paper II .....	28
Paper III.....	28
Paper IV.....	29
<b>PATIENTS AND METHODS.....</b>	<b>30</b>
<i>TUMOURS, CELL LINES AND CONTROL MATERIAL.....</i>	<i>30</i>
Paper I.....	30
Paper II .....	30
Paper III.....	30
Paper IV.....	30
<i>METHODS.....</i>	<i>31</i>
Polymerase Chain Reaction (PCR).....	31
Mutation screening methods.....	32
Sanger sequencing.....	32
Principles of Massively Parallel Sequencing (MPS).....	33
Bridge amplification and sequencing by synthesis (Illumina).....	33
Ion Torrent Personal Genome Machine (PGM™) technology .....	35
HaloPlex™ Target Enrichment System.....	36
Data processing.....	37
CNV methods.....	39
Single Nucleotide Polymorphism (SNP) array .....	39
Data Analysis & software.....	40
Cell culture and immunoblotting.....	41
Statistical methods .....	41
Fisher's exact test.....	41
Kaplan-Meier survival analysis.....	41
<b>RESULTS AND DISCUSSION .....</b>	<b>43</b>
Paper I.....	43
Paper II .....	44
Paper III.....	45
Paper IV.....	47
<b>CONCLUSION.....</b>	<b>50</b>
<b>FUTURE PERSPECTIVE .....</b>	<b>51</b>
<b>ACKNOWLEDGEMENT .....</b>	<b>52</b>
<b>LITERATURE CITED.....</b>	<b>54</b>



# ABBREVIATIONS

ALCL	anaplastic large cell lymphoma
ALK	anaplastic lymphoma kinase
ALKAL1/2	ALK and LTK ligand 1/2 (FAM150A/FAM150B)
bp	base pair
BTD	breakthrough therapy designation
CNV	copy number variants
CpG	cytosine guanine dinucleotide
DM	double minute
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferases
ddNTP	dideoxynucleotides
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
ECD	extracellular domain
FAM150	family with sequence similarity 150
FDA	food and drug administration
HSR	homogenously staining regions
INRGSS	neuroblastoma risk group staging system
INSS	International neuroblastoma staging system
kb	kilobasepairs, thousands of base pairs
LOH	loss of heterozygosity
lncRNA	long non-coding RNAs
MAF	mutated allele fraction
Mb	megabasepairs, millions of base pairs

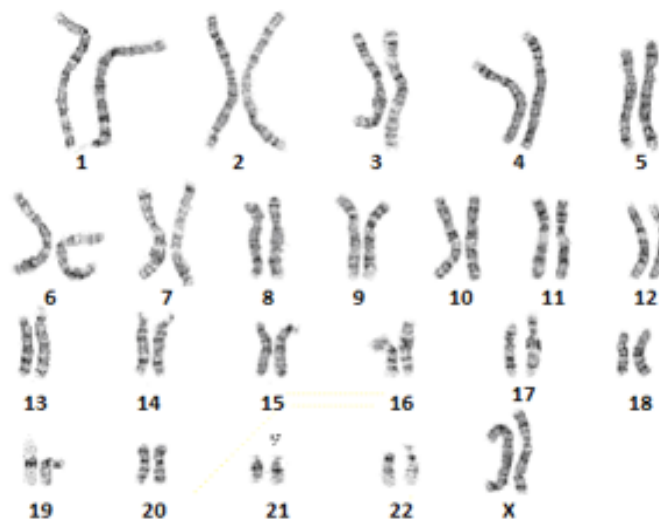
MPS	massively parallel sequencing
miRNA	microRNAs
mRNA	messenger RNA
NB	neuroblastoma
ncRNA	non-coding RNA
NGS	next generation sequencing
NPM	nucleophosmin
NSCLC	non-small cell lung cancer
PCR	polymerase chain reaction
piRNA	PIWI-interacting RNA
RNA	ribonucleic acid
RTK	receptor tyrosine kinase
SCA	segmental chromosome alterations
TAD	topologically associated domain
siRNA	short interfering RNA
SNP	single nucleotide polymorphism
SNS	sympathetic nervous system
SRO	smallest region of overlap
TKD	tyrosine kinase domain
TKI	tyrosine kinase inhibitor
tRNA	transfer RNA
TSG	tumour suppressor gene
UCSC	University of California, Santa Cruz
UTR	untranslated regions

# INTRODUCTION

## BASIC GENETICS

### DNA and genes

In human cell nuclei, the genetic information is stored into two sets of 23 DNA molecules called chromosomes, half of which are inherited from the mother and half from the father. In eukaryotes, DNA is found in the nucleus, in the mitochondria and also in chloroplast of plants. A DNA molecule consists of two twisted complementary paired strands, called helix, as described by Watson and Crick in 1953 (Watson & Crick 1953). Each strand is built up by four different nucleotides bases, adenine (A), cytosine (C), guanine (G) and thymine (T). Each nucleotide in one chain is specifically linked to a nucleotide in other chain referred to as base pairs (bp), an A always pair with a T and a C with a G. The usage of different combinations of these bases makes the genetic code.



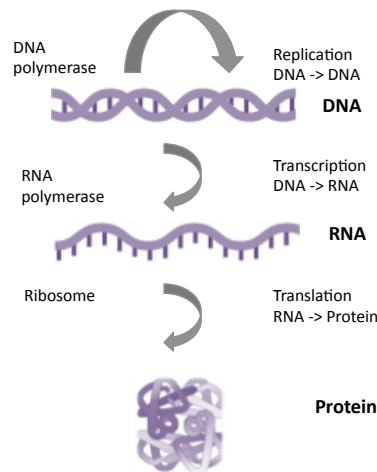
**Figure 1** A human female karyotype showing 46 chromosomes, 23 chromosome pairs.

The human genome contains approximately 3.2 billion nucleotides and about 19,000 protein coding genes (Ezkurdia et al 2014). Each gene is comprised of the

protein coding segments, known as exons; the intervening non-coding sequences, known as introns; and the regulatory regions on each end of the gene (5' and 3' end regions). The first and the last exon commonly also include untranslated regions (UTRs), which is important for RNA stability and translation. There are about 180,000 exons in each human genome that are collectively referred to as an exome. An exome comprises about 1% of the human genome and hence, is about 30 million nucleotides in size.

### The central dogma of molecular biology

The flow of genetic information from DNA to RNA to polypeptide has been described as the central dogma (Crick 1958). During transcription, the RNA-polymerase transcribes the DNA gene sequence to a newly synthesized pre-mRNA (messenger RNA). The pre-mRNA is processed to a mature RNA through excision of introns, addition of a 5' methylated nucleoside as protective cap and polyadenylation at the 3' end. Following post-transcriptional modification, the mRNA is released from the DNA and will be transported out from the nucleus to the cytoplasm to be used as a blueprint for the protein to be produced. After migration to the cytoplasm, the mRNA is translated into polypeptides at the ribosome (a multiunit RNA-protein complex). The connection between mRNA and protein is made through a set of transfer RNAs (tRNAs), each containing a three nucleotides long recognition sequence. To become a fully functional protein the amino acid chain folds up into a unique three-dimensional configuration and possibly undergoes post-translational modification.



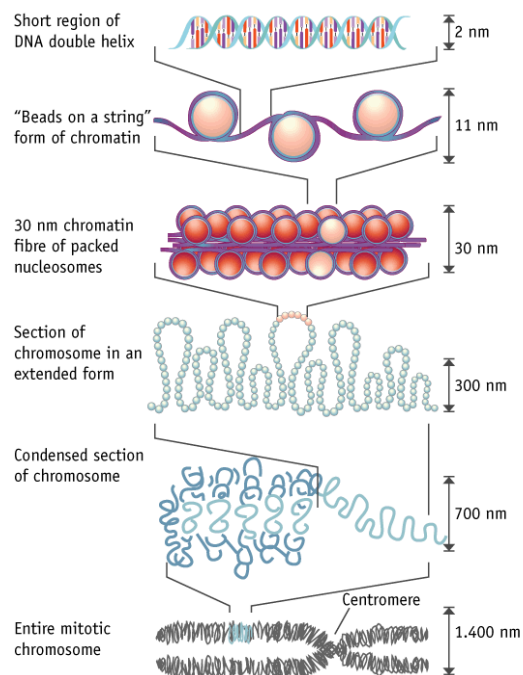
**Figure 2.** The central dogma of molecular biology.

## Genetic variations

A single nucleotide polymorphism (SNP),  $3.5 \times 10^6$  abundance in the human genome, is a common variation in the DNA sequence in which one nucleotide at the specific locus differs between individuals. Usually, a SNP has two alleles, although three and four allele SNPs exist, but they are rare. There is also a form of large scale polymorphism that involves DNA copy number variants (CNVs), defined as stretches of DNA larger than 1 kb that display copy number differences in the normal population (Scherer et al 2007). Change of copy number is implicated to be important in functional diversity and individual CNVs have shown to be associated with disease or susceptibility to disease (de Smith et al 2008). Polymorphism definition refers to a DNA sequence variation that occur naturally in population and are not disease causing. However, the term “mutation” is more commonly used for pathogenic alteration in DNA. Mutations result from unrepaired damage to DNA, errors in the process of replication, or from the insertion or deletion of segments of DNA by mobile genetic elements. Mutations are often found in coding regions of the gene, but could also be found in regulatory elements and other locations that could have implications in both normal and abnormal biological processes including evolution, cancer, and the development of the immune system. The sequence of a gene can be altered in a number of ways. Point mutations, exchange of a single nucleotide for another, which results in an amino acid variation, is called nonsynonymous, and one that does not is called synonymous. Nonsynonymous mutations can be further divided into missense and nonsense variations. A missense variation results in a different amino acid and a nonsense variation results in a premature stop codon that causes a truncated protein. In addition to basepair substitution, mutations can also result from deletions or insertions in the coding region of a gene, which may alter splicing of the mRNA (splice site mutation), or cause a shift in the open reading frame if the number of the involved nucleotides is not a multiple of three (frameshift mutation). Cancer cells often exhibit large-scale chromosomal abnormalities, including deletions, amplifications, translocations (interchange of genetic parts from nonhomologous chromosomes) and inversion.

## Organization of the genetic material

Each human cell has about 2 meters of DNA, which is organized by the core-packing unit of chromosome called nucleosome. The nucleosomes core particle consists of approximately 147 base pairs of DNA coiled around a histone octamer (two each of H2A, H2B, H3, and H4). The adjacent nucleosomes are connected by stretches of "linker DNA", viewed as “beads on a string” in an electron microscope. This “beads on a string” is compacted into a chromatin fibre that is condensed into long loops. Further compaction leads to transcriptionally inactive heterochromatin that can be visualized as chromosomes in light microscope during cell division. Between divisions, nucleosomes are relatively loosely packaged allowing expression of genetic information.



**Figure 3** The organization of DNA within the chromatin structure. Reprinted with permission from Macmillan Publishers Ltd: *Nature* (Felsenfeld & Groudine, 2003), © 2003

## EPIGENETICS

Epigenetic term refers to the cellular and physiological heritable traits that are caused by mechanisms other than changes in underlying DNA sequence, a change in phenotype without a change in genotype. Epigenetic is important for X chromosome inactivation in female mammals and is also responsible for chromosome imprinting. At least three systems, including DNA methylation, histone modification and non-coding RNA (ncRNA)-associated gene silencing are currently considered to initiate and sustain epigenetic change (Egger et al 2004).

### DNA methylation

DNA methylation is one of the most well characterized epigenetic modifications dating back to studies done by Griffith and Mahler in 1969 which suggested that DNA methylation may be important in long term memory function (Griffith & Mahler 1969; Holliday 2006). During methylation, DNA methyltransferases

(DNMTs) enzyme adds a methyl group to cytosine in CpG dinucleotides (cytosine - guanine dinucleotides). The majorities of CpG sites are located within repetitive elements and are methylated. Another place where they are found is in CpG islands associates with the promoter of genes, normally unmethylated (Esteller 2002). Adding methyl groups modifies a gene's interactions with the machinery within a cell's nucleus that is needed for transcription and is associated with gene silencing.

## Histone Modifications

When histones are modified after they are translated into protein, i.e. post-translation modification, they can influence how chromatin is arranged, which in turn can determine whether the associated chromosomal DNA will be transcribed. Histones can be modified in different ways including methylation, acetylation, phosphorylation, and ubiquitination. Acetylation is usually associated with active chromatin, while deacetylation is generally associated with heterochromatin. It has been proposed that the combination of modifications constitute a so called "histone code" which defines the status of the chromatin structure (Jenuwein & Allis 2001).

## RNA-associated silencing

Small, non-coding RNAs that bind to target mRNAs transcripts in a sequence-specific manner can also turn off genes activity. The main classes of small RNA are short interfering RNAs (siRNA), microRNAs (miRNA) and PIWI-interacting RNAs (piRNAs) (Jinek & Doudna 2009). The siRNA and miRNA origin from double-stranded RNA (dsRNA) and their generation depends on ribonuclease (RNase) Dicer. The siRNA has a double-stranded structure and the miRNA a single-stranded. Little is known about the piRNA, but they are generated from single-stranded RNA and have a function in silencing transposons in germ cells. Moreover, long non-coding RNAs (lncRNA) have recently emerged as gene regulators in several cancers.

## TAD regulation in the 3D genome

Three-dimensional organization of the genome occurs coincident with genetic and epigenetic alterations. Structural boundaries between regions of chromatin define loops called topologically associated domains (TADs), within which gene activity is coordinated (Dixon et al 2012). TADs are conserved during evolution and play roles in controlling long-range chromatin interactions. They have contributed to the increment of regulatory complexity by allowing promoters to be activated or inactivated with different sets of cis-regulatory elements.

# CANCER

Cancer is a genetic disease characterized by out-of-control cell growth. Cancer is the result of several somatically acquired mutations and occasionally also an inherited predisposition. Cancer is not one single disease, there are over a hundred different types of cancer. The progression and pathology varies between types of cancers and between individuals and fighting the disease has therefore proven to be very challenging. Sweden has about 50,000 new cases each year and approximately 22,000 dies from cancer annually.

## Oncogenes and tumour suppressor genes

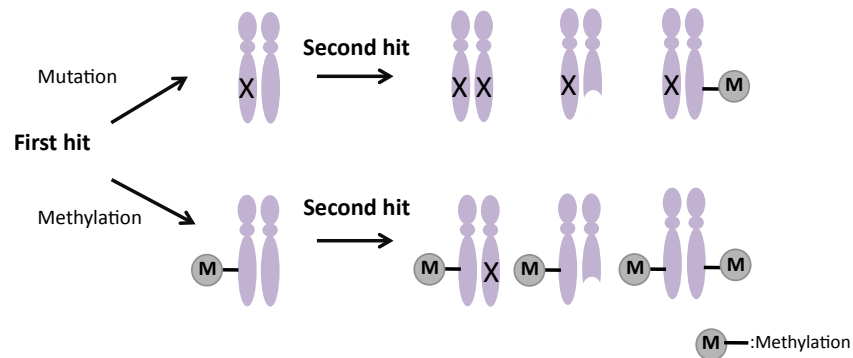
Two major groups of genes are involved in tumourigenesis, oncogenes and tumour suppressor genes (TSG). Oncogenes are altered versions of normal proto-oncogenes whose normal function promotes cell proliferation. The gain of function mutations in these genes result in an excessive or inappropriate activation.

Tumour suppressor genes (TSG) are inhibiting uncontrolled cell growth and inactivation of TSGs is just as important as oncogene activation in cancer pathogenesis. TSGs can be divided into gatekeepers and caretakers (Kinzler & Vogelstein 1997). Gatekeepers are genes that directly regulate the growth of tumours by inhibiting proliferation or promoting apoptosis. Both the maternal and paternal copy of the gene need to be altered for a tumour to develop and the inactivation of the gatekeepers is therefore rate limiting for the initiation of a tumour. The inactivation of a caretaker gene does not promote tumour initiation directly but leads to increased genetic instability that in turn leads to mutations of other genes.

## The two-hit hypothesis

Knudson's two-hit hypothesis states that two hits are needed for a TSG to be inactivated, exemplified by a childhood eye tumour, retinoblastoma (Knudson 1971). The first hit can be a de novo mutation as in sporadic cancers or germline as in familiar cancer syndromes. The second hit can constitute of a second mutation, but more often a deletion of the remaining allele or promoter hypermethylation that silences this allele (Jones & Laird 1999), see Figure 4. The two-hit hypothesis is however not universal and there is evidence that haploinsufficiency, inactivation of one allele, can contribute to the tumourigenic process.





**Figure 4.** Mechanisms of TSG inactivation in induction of cancer. (A) The first hit is a mutation affecting function of TSG. (B) The first hit is DNA hypermethylation which silences the gene. The second hit commonly constitutes somatic cell mutation, heterozygote loss, and DNA hypermethylation.

## Epigenetic and cancer

Cancer was the first human disease to be linked to epigenetics (Feinberg & Vogelstein 1983). DNA hypomethylation can activate oncogenes and initiate chromosome instability, whereas DNA hypermethylation initiates silencing of tumour suppressor genes (Sakai et al 1991). An accumulation of genetic and epigenetic errors can transform a normal cell into an invasive or metastatic tumour cell. Subsequently, epigenetic changes can be used as biomarkers for the molecular diagnosis of early cancer.

## TAD and cancer

Disruption of TAD boundaries, as recently detected in cancer cells, can lead to inappropriate promoter-enhancer communication and misregulation of oncogenes and tumour suppressors (Flavahan et al 2016; Valton & Dekker 2016). This alteration in the 3D folding of chromosomes contributes to oncogenesis through two mechanisms. One mechanism locally disrupts domains by mutating or epigenetically inactivating a TAD boundary. However, the other mechanism involves genomic rearrangements that break up TADs and creates new ones without directly affecting TAD boundaries.

# NEUROBLASTOMA

## Epidemiology

Neuroblastoma (NB) is the most frequent extracranial solid tumour in children, accounting for 7 to 8% of all childhood malignancies and 15% of all cancer-related deaths in this population. It is the most frequently diagnosed neoplasm during infancy (Brodeur 2003; Maris & Matthay 1999). The prevalence is about 1 in 7,000 live births, with 15-20 new diagnosed cases a year in Sweden. The median age at diagnosis is approximately 18 months, 9 months in familial neuroblastoma. While 90% of the patients are younger than 5 years, NB is very rare after the age of 10 (Brodeur 2003). It is an embryonic tumour of the postganglionic sympathetic nervous system (SNS). Most NBs are composed of neuroblasts, undifferentiated sympathetic nerve cells arising from the neural crest. Tumours characterized by partial histological differentiation are called ganglioneuroblastoma (GNB) while the most differentiated tumour type is called ganglioneuroma (GN). Primary tumours are found in areas of the peripheral SNS, about half of all NBs arise in one of the adrenal glands, and the rest develop in paraspinal sympathetic ganglia in the chest, abdomen, or pelvis. Metastases often spread to regional lymph nodes, bone, bone marrow, and sometimes to the liver and skin in infants. NB is remarkable for its clinical heterogeneity. The clinical course can vary enormously ranging from spontaneous regression, particularly observed in infants, to inexorable progression and death despite multimodal treatments. Metastatic disease is present in approximately 50% of cases.

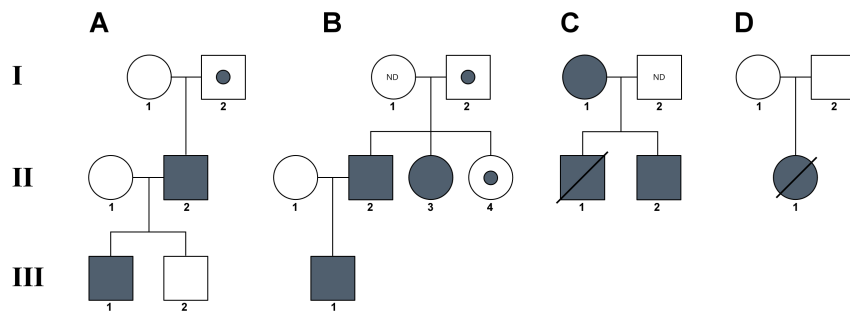
## Symptoms and treatment

The symptoms of neuroblastoma are usually diffuse depending on the size and location of the primary tumour, the extent of spread to other parts of the body and whether or not the tumour cells secrete hormones. Weakness, fever, high blood pressure, pain, and weight loss are some of the symptoms while in some other patients it can remain asymptomatic. Due to the extreme heterogeneity in the clinical behaviour of this disease, a challenging task is not only to find a cure for the high-risk patients, but also to avoid overtreatment for the patients with a favourable prognosis. Surgical removal, chemotherapy, radiotherapy and biotherapy are among the treatment strategies for NB patients.

## Hereditary neuroblastoma

Hereditary neuroblastoma, which segregates as an autosomal dominant Mendelian trait with incomplete penetrance, is rare (Knudson & Strong 1972; Kushner et al

1986; Liu & Thiele 2012). Patients with familial neuroblastoma often have multiple primary tumours and are characterized by an earlier median age at diagnosis. Similar to sporadic NB patients, hereditary cases also show striking heterogeneity in the type of tumours that arise. Missense or nonsense germline mutations in *PHOX2B*, a homeobox gene that is a master regulator of normal autonomic nervous system development, explain a small subset of hereditary neuroblastoma, as well as other disorders of neural crest-derived tissues (Bourdeaut et al 2005; Krona et al 2008; Mosse et al 2004; Trochet et al 2004). However, activating mutations in the tyrosine kinase domain of anaplastic lymphoma kinase gene (*ALK*) located on the short arm of chromosome 2 seem to explain the majority of hereditary neuroblastomas as well as sporadically occurring cases particularly in high stage tumours. (Janoueix-Lerosey et al 2008; Mosse et al 2008).



**Figure 5.** Pedigrees representing four different NB families. The panels show the family members *ALK* status, and if they have been diagnosed with NB (filled symbols). Squares indicate male family members and circles females. A diagonal line marks a diseased individual. Symbols with a filled dot inside represent *ALK* mutation carriers, free of disease. A family member where *ALK* analysis was not done is marked with ND (unpublished data).

## Prognostic factors

The likelihood of cure depends on several prognostic factors, including stage of disease, age at diagnosis, histological features, and several genetic and biological markers. Children less than one year of age generally have a much better prognosis than children diagnosed above this age, with equivalent stages (Breslow & McCann 1971).

### *Expression of neurotrophin receptors*

The interaction between neurotrophic factors and tyrosine kinase receptors has an essential role in nerve cells development as well as neuroblastoma behaviour. TrkA expressing neuroblastomas have favourable prognosis as a result of mediating

apoptosis or differentiation (Kogner et al 1993; Nakagawara et al 1992; Ryden et al 1996; Suzuki et al 1993). TrkC is also involved in low stage neuroblastoma in absence of *MYCN* amplification (Ryden et al 1996; Yamashiro et al 1996). However, expression of full length TrkB is correlated with *MYCN* amplification and aggressive tumours (Nakagawara et al 1994). A truncated form of TrkB is mainly found in low stage tumours.

### *Ploidy*

The overall copy number or ploidy of neuroblastoma tumours has been found to be of prognostic importance (Cohn et al 2009). A hyperploid or near-triploid karyotype is representative of favourable NB whereas those with near-diploid or near-tetraploid DNA content show metastatic behaviour and poor prognosis (Kaneko et al 1987; Ladenstein et al 2001). It is assumed that near-triploid cases have a fundamental defect in mitosis leading to gains and losses of whole chromosomes, or numerical aberration. Near-diploid or near-tetraploid karyotypes have a fundamental defect in genomic stability leading to chromosomal rearrangements such as unbalanced translocations, or segmental aberrations (Kaneko et al 1987).

### *Tumour histology*

To classify NBs for prognostic use according to their histological features, the Shimada histopathologic grading system was devised in 1984. It is an age-linked classification system that divides the tumours into favourable and unfavourable histology groups depending on the degree of neuroblast differentiation, schwannian stroma content, and mitosis-karyorrhexis index. A high mitosis-karyorrhexis index (MKI) in stroma poor tumours is an indicator of poor prognosis in Shimada system.

### *Risk stratification*

The International Neuroblastoma Staging System (INSS) was developed in 1986 (and revised in 1993) to facilitate the comparison of clinical trials worldwide by taking into account the age at diagnosis, the spread of disease, and the respectability of the tumour, grading NBs from stage 1 to 4, with stage 4 being the most aggressive tumours, see Table 1 (Brodeur et al 1993; Brodeur et al 1988).

Stage	Definition
1	Localized tumour with complete gross excision, with or without microscopic residual disease; representative ipsilateral lymph nodes negative for tumour microscopically
2A	Localized tumour with incomplete gross excision; representative ipsilateral lymph nodes negative for tumour microscopically
2B	Localized tumour with or without complete gross excision, with ipsilateral lymph nodes positive for tumour; enlarged contralateral lymph nodes must be negative microscopically
3	Unresectable unilateral tumour infiltrating across the midline, with or without regional lymph node involvement; localized unilateral tumour with contralateral regional lymph node involvement; or midline tumour with bilateral extension by infiltration (unresectable) or lymph node involvement
4	Any primary tumour with dissemination to distant lymph nodes, bone, bone marrow, liver, and other organs (except as defined for stage 4S)
4S	Localized primary tumour (as defined as stage 1, 2A, or 2B), in patient <1 year, with dissemination limited to skin, liver, or bone marrow (marrow involvement should be minimal with malignant cells <10% of total nucleated cells)

**Table 1.** *International Neuroblastoma Staging System. Adopted from (Brodeur et al 1993)*

The INSS has been widely used for the last 20 years for the treatment stratification of neuroblastoma tumours. However, since the INSS relies on excision of the primary tumour to assign patients to stage 1 and 2, it is not suitable as a pretreatment staging system and risk assessment. As a consequence, the International Neuroblastoma Risk Group Staging System (INRGSS) has recently been established (Monclair et al 2009). It is based on tumour imaging rather than the extent of surgical resection, Table 2 & 3. The International Neuroblastoma Risk Group (INRG) classification system includes INRGSS, age, histology, and grade of tumour differentiation, *MYCN* status, 11q deletion status, and tumour cell ploidy to classify NB tumours.

Stage	Description
L1	Localized tumour not involving vital structures as defined by the list of image-defined risk factors and confined to one body compartment
L2	Locoregional tumour with presence of one or more image-defined risk factors
M	Distant metastatic disease (except stage MS)
MS	Metastatic disease in children younger than 18 months with metastases confined to skin, liver, and/or bone marrow (< 10%)

**Table 2.** *International Neuroblastoma Risk Group (INRG) Staging System. For image-defined risk factors see (Monclair et al 2009)*

INRG stage	Age (months)	Histologic category	Grade of tumor differentiation	MYCN	11q-del	Ploidy	Pretreatment Risk Group
L1/L2		GN maturing; GNB intermixed					A Very low
L1		Any, except GN maturing or GNB intermixed		NA Amp			B Very low K High
L2	<18	Any, except GN maturing or GNB		NA	No		D Low
			Differentiating	NA	Yes		G Intermediate
	≥18	GNB nodular; Neuroblastoma			No		E Low
			Poorly differentiated or undifferentiated	NA	Yes		H Intermediate
M	<18			Amp NA		Hyper-diploid	N High F Low
	<12			NA		Diploid	I Intermediate
	12 to <18			NA		Diploid	J Intermediate
	<18			Amp			O High
	≥18						P High
MS	<18			NA	No		C Very low
				Amp	Yes		Q High R High

**Table 3.** INRG pretreatment risk classification. Adopted from (Cohn et al 2009; Monclair et al 2009)  
GN = ganglioneuroma; GNB = ganglioneuroblastoma; Amp = amplified; NA = not amplified

## Chromosomal abnormalities

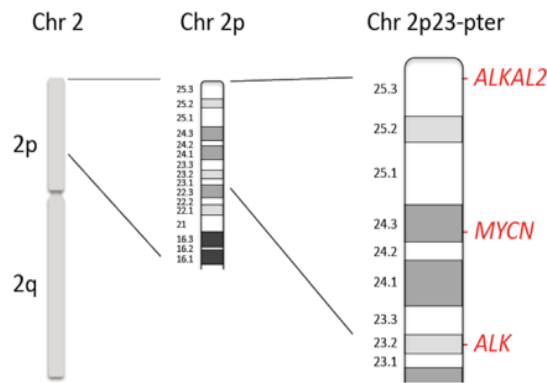
### *1p deletion*

Deletion of the short arm of chromosome 1 (1p36.3), first reported by (Brodeur et al 1977), occurs in 25-35% of all NB (Bauer et al 2001; Caren et al 2008b; Maris et al 2001; Martinsson et al 1995). This aberration is correlated with *MYCN* amplification and is found in approximately 70% of high-risk NBs (Maris et al 2000). The deletion is mainly large and many investigations have been made to identify the shortest region of overlap (SRO) of deletions in this region. However, the identified regions are not entirely consistent, indicating this region harbours several tumour suppressor genes. Whether the loss of heterozygosity (LOH) due to deletion of alleles from 1p is an independent indicator of prognosis, remains controversial.

*Chromosome arm 2p: MYCN, ALK, ALKAL2*

Chromosome region 2p24 amplification is detected in 15-30% of NB tumours and is associated with advanced disease stage and malignant progression (Brodeur et al 1984; Caren et al 2008b; Schwab et al 1983).

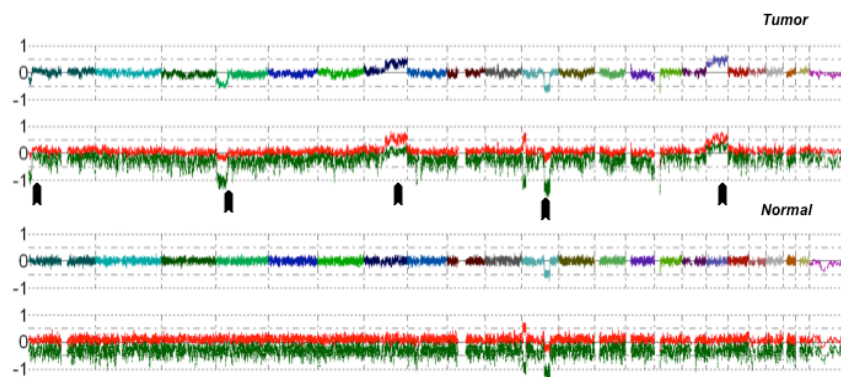
The size of amplified DNA encompassing *MYCN* proto-oncogene can range from 100 Kb to more than 1 Mb which raises the possibility that additional target genes may be co-amplified. However, *MYCN* (>>100 copies per cell) has emerged as the only gene consistently present when DNA is amplified in neuroblastoma cells. Homogenously staining regions (HSRs) and double minutes (DMs) chromatin bodies were found as the chromosomal sites of amplified *MYCN*. DMs are located extra-chromosomally, whereas HSRs are located at different chromosomes and usually not at the resident site. *MYCN* encodes for a transcription factor that is normally expressed during embryonic development. Another important player in neuroblastoma, *ALK* gene is also situated on chromosomal region (2p23.2), often present in 2p gain region found in primary NB. Moreover, the *ALK* gene is identified to be mutated or amplified in a majority of familial cases and is also affected in sporadic NB (Caren et al 2008a; Chen et al 2008; Janoueix-Lerosey et al 2008; Mosse et al 2008). Activating point mutation in tyrosin kinase domain of *ALK* is found in 6-12% of sporadic NB cases and *ALK* gene amplification, frequently cooccurring in the *MYCN* amplicon, in 3-5%. Furthermore, the recently described *ALK* ligand *ALKAL2*, can also be found within this 30Mb distal portion of chromosome 2p and may play a role in wildtype *ALK* activation, see Figure 6 (Guan et al 2015; Reshetnyak et al 2015).



**Figure 6.** *ALKAL2, ALK and MYCN on 2p. Schematic overview of human chromosome 2 showing localization of ALK, MYCN and ALKAL2 in the 2p23-pter region. ALK is located at 2p23.2, MYCN at 2p24.3 and ALKAL2 more distally at 2p25.3.*

### 11q loss

Deletion of the long arm of chromosome 11 (11q23) is found in 20-45% of NB tumours (Attiey et al 2005; Caren et al 2008b; Srivatsan et al 1993). It is inversely correlated with *MYCN* amplification and is a good predictor of aggressive tumours that lack *MYCN* amplification (Guo et al 1999). The high frequency of chromosomal breaks in the unfavorable 11q deletion group of tumours is suggestive of a chromosomal instability phenotype gene located in 11q and involved in DNA repair system.



**Figure 7.** Rare case of a constitutional deletion involving the long arm of chromosome 11. The multi-coloured line shows the total copy number for the different chromosome regions and the red and green line show the strongest and weakest allele intensity respectively. We observed the identical 11q del (81.5-102.9 Mb) in tumour to that found in constitutional DNA and additionally del 1p (pter-12.1 Mb); del 4p (4.1-33.2 Mb); dup 7q (76.8 Mb-qter); dup 17 alterations have been observed (indicated by arrows). del, deletion; dup, duplication. Adopted from (Passariello et al 2013)

### 17q gain

The gain of parts of the long arm of chromosome 17 (17q25) has been detected in about 50% of NB tumours and known as the most frequent genetic abnormality of tumours (Abel et al 1999; Caren et al 2008b; Cohn et al 2009; Gilbert et al 1984). Gain of 17q always involves the terminal part of 17q. It is hypothesized that the gain of genomic material at terminal of 17q can possibly result in a growth advantage for tumour cells (Schleiermacher et al 2004). Unbalanced t(1;17) translocation resulting in 1p deletion and 17q gain also occurs frequently (Savelyeva et al 1994).



### *Other chromosomal aberrations*

Other regions deleted in neuroblastoma include allelic loss for 14q (14q23-32), and 3p (3p22), which are present in 20% and 15% of NB respectively (Hallstenson et al 1997; Srivatsan et al 1993; Suzuki et al 1989). Similar to 11q, 14q deletion group also show an inverse correlation with *MYCN* amplification. However, LOH for 14q was evident in all clinical risk groups, suggesting that this rearrangement may be a universal feature of neuroblastoma tumour development (Thompson et al 2001).

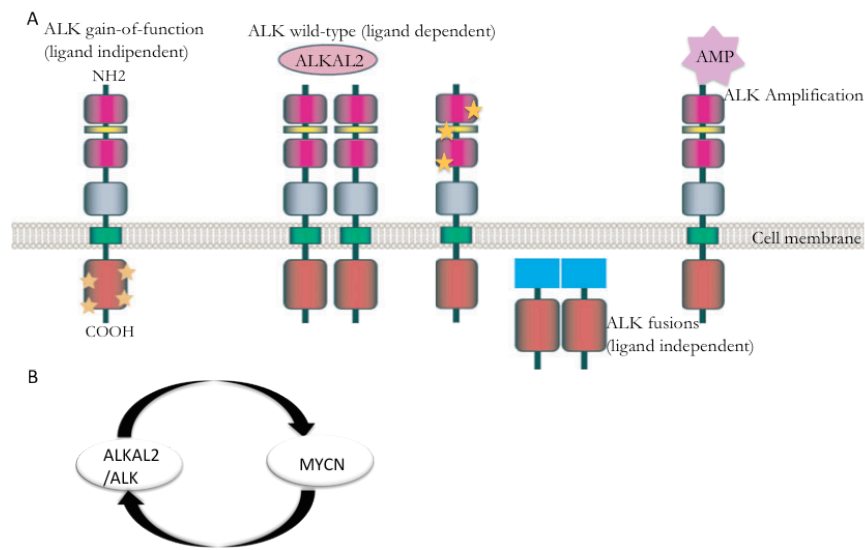
### *ALK (anaplastic lymphoma kinase)*

In addition to major chromosomal rearrangements, there are recurrent genetic alterations in the *ALK* gene in NB. *ALK* is a receptor tyrosine kinase (RTK) encoded by the *ALK* gene located at chromosome 2p23. *ALK* was first identified in 1994 as part of a fusion partner of *NPM* (nucleophosmin) in the t(2,5)(p23,q35) translocation in anaplastic large cell lymphoma (ALCL) (Morris et al 1994; Shiota et al 1994). Indeed, *ALK* fusion proteins due to chromosomal translocations have been observed in non-small cell lung cancer and inflammatory myofibroblastic tumours but is a rare event in NB (Fransson et al 2015; Kwak et al 2010; Pulford et al 2004). Subsequent studies have shown that various fusion genes involving *ALK* together with gene amplification or single nucleotide mutations are common mechanisms leading to increased *ALK* activity (Pulford et al 1997). The full length 220kDa *ALK* protein contains three domains, an extracellular ligand binding domain, a transmembrane region and an intracellular tyrosine kinase domain (TKD). The kinase domain contains a motif with three tyrosine's YxxYY (Tyr1278, Tyr1282, and Tyr1283) that are the major autophosphorylation sites regulating *ALK* activity. In situ hybridization in mouse showed that *ALK* expression was restricted to specific areas in the developing brain with a decrease to barely detectable levels after birth (Iwahara et al 1997). Mammalian *ALK* plays a central role in normal developing of the nervous system however when deregulated and constantly activated *ALK* has a great oncogenic potential.

### *ALK, MYCN and NB*

In 2000 Lamant et al found full length *ALK* expression in NB (Lamant et al 2000). The *ALK* gene was further reported to be amplified in NB cell lines and tumours (Miyake et al 2002; Osajima-Hakomori et al 2005). In 2008 Mosse et al identified *ALK* as a major familial NB predisposition gene, showing that germ line mutations in *ALK* explain most familial neuroblastoma cases. The Mosse group and also our group could furthermore show that sporadic mutations in *ALK* kinase domain as well as *ALK* amplification were frequent in high-risk tumours (Caren et al 2008a; Mosse et al 2008). In addition, *ALK* can also be activated by rare translocation events in NB (Cazes et al 2013; Fransson et al 2015). *ALK* mutations are present

across all neuroblastoma risk groups but has a notable contribution in *MYCN*-amplified tumours where the combination of these aberrations result in a particular poor outcome (Bresler et al 2014; De Brouwer et al 2010). The stimulated wildtype as well as ligand-independent gain-of-function mutants in *ALK* stimulate initiation of transcription of the *MYCN* which, in turn, directly regulates *ALK* expression (Hasan et al 2013; Schonherr et al 2012). Finally, cotransfection of *ALK* gain-of-function mutations and *MYCN* causes increased transforming capacity leading to earlier onset and higher penetrance possibly due to blocking of *MYCN*-driven apoptosis (Berry et al 2012; Schonherr et al 2012; Zhu et al 2012). The *ALK*-driven neuroblastoma mouse model recapitulates syntenic chromosomal aberrations, including *MYCN* amplification, to human neuroblastoma suggesting a synergism between *MYCN* and *ALK* (Heukamp et al 2012).



**Figure 8.** The *ALK* receptor tyrosine kinase. (A) different mechanisms of *ALK* signalling activation in cancer. The full-length *ALK* receptor comprising an amino-terminal ECD and an intracellular tyrosine kinase domain (in red), connected by a single transmembrane domain (in green). The *ALK* ECD contains two MAM domains (in pink), one LDLA domain (in yellow) and a glycine-rich region (in grey). From right to left: activation (gold stars) by point mutations in the context of full-length *ALK*, activation by the small secreted ALKAL2 ligand, activation by intragenic translocation of *ALK* and subsequent partial loss of ECD, activation by fusion partner (in blue) fused to the intracellular tyrosine kinase domain of *ALK* (in red) (not seen in NB), activation by *ALK* amplification. (B) Simple model illustrating a putative positive feedback loop in ALKAL2-*ALK*-*MYCN* signalling events in neuroblastoma. Adopted from (Hallberg & Palmer 2016)

### *ALK as a therapeutic target*

Incidence of relapse is known as the major reason for treatment failure in high-risk NB patients (Sung 2012). To date, in neuroblastoma, only limited data is available regarding genetic alterations at the time of tumour progression or relapse. An accumulation of segmental chromosome alterations at the time of relapse has been described (Schleiermacher et al 2010), including tumours of patients who did not receive any DNA damaging treatment. The possibility of occurrence of new *ALK* mutations at the time of progression or relapse is of high clinical importance as this may lead to new therapeutic possibilities using tyrosine kinase inhibitors (Schleiermacher et al 2014). Moreover, The findings on *MYCN* and *ALK* cooperation in NB suggest that *ALK* inhibitors also can be used in treatment regime of *MYCN*-amplified NB patients regardless of *ALK* mutation status – obviously a very exciting finding (Schonherr et al 2012; Zhu et al 2012).

### *Tyrosine kinase inhibitors*

Due to the importance of activated signal transduction in tumorigenesis a wide array of different tyrosine kinase inhibitors has been developed (Druker et al 1996; Li & Morris 2008). The *ALK* mutated variants confer not only differential oncogenic activity but also show differential sensitivity to *ALK* inhibition therapy (Bresler et al 2014). The first clinically approved inhibitor of *ALK*, *c-MET* and *ROS1*, crizotinib (FDA approved AUG 2011), has shown remarkable clinical efficacy in treatment of NSCLC and ALCL harbouring *ALK* fusions (Christensen et al 2007; Mosse et al 2013). However, cellular responses to mutant *ALK* are complex when compared to rearranged *ALK*, and treatment remains a challenge (Bosse & Maris 2016; Hallberg & Palmer 2013). In neuroblastoma, activating mutations in the *ALK* kinase domain are typically refractory to crizotinib treatment, highlighting the need for more potent inhibitors. Recently, second generation *ALK* inhibitors have been developed and approved for use in patients with NSCLC. Alectinib (CH5424802) (Lu et al 2017) and brigatinib (AP26113) (Katayama et al 2011) received FDA granted breakthrough therapy designation (BTD) in 2013 and 2014 respectively. Brigatinib is able to overcome crizotinib-induced chemoresistance and effectively induces apoptosis in human neuroblastoma (Siaw et al 2016). Ceritinib (LDK378; FDA approved 2014) is also approved for use in crizotinib-relapsed NSCLC patients in the US and Europe (Marsilje et al 2013; Shaw et al 2014). Lorlatinib (PF-06463922) is a third-generation *ALK*-TKI currently progressing in phase III clinical trials in NSCLC patients (Basit et al 2017; Guan et al 2016; Infarinato et al 2016; Janoueix-Lerosey et al 2008). While the above-mentioned drugs are all ATP-competitive inhibitors of *ALK*, they have differential binding affinity and therefore differential efficiency in blocking the activity of the various *ALK* resistant mutant forms (Hallberg & Palmer 2013; Mullis et al 1986). Thus, a complex picture of *ALK* inhibition is emerging, with multiple distinct resistant subclones arising following primary treatment with particular *ALK* inhibitors.

# OBJECTIVES

The overall aim of this thesis was to delineate the culprit genes that may play a critical role in development and/or progression of neuroblastoma and also find new targets that can be used in patient stratification and precision medicine.

## *Paper I*

- To retrospectively determine the frequency by which the therapeutically relevant *ALK* hotspot mutations is discovered at both diagnosis and relapse time points
- To evaluate the role of activating point mutations in the TKD of *ALK* in neuroblastoma progression or relapse and to define its potential role in clonal evolution of neuroblastoma
- To infer if the serial biopsy procedures followed by deep *ALK* sequencing can provide the potential for patient benefit highlighting the usefulness of targeted therapy with ALK inhibitors

## *Paper II*

- To put forward the hypothesis that the gain of the distal region of chromosome 2p, interestingly harbouring three key players in the ALK signalling pathway, namely MYCN, ALK and the ALK ligand ALKAL2, has important implications for the development of NB. This prompted us to investigate this connection in tumours from NB patients

## *Paper III*

- To explore the variation and frequency of *ALK* hotspot mutations that can go undetected using Sanger-based sequencing chemistry
- To further define the subclonal heterogeneity of the tumour and their potential role in clonal evolution and progression of neuroblastoma

*Paper IV*

- To characterize the genomic structural variants at basepair resolution, and define breakpoint mapping and SRO of deletions and gain, to arrive at novel candidate genes in 11q-deleted neuroblastoma tumours using genome-wide approach

# PATIENTS AND METHODS

## TUMOURS, CELL LINES AND CONTROL MATERIAL

In this thesis, we have used primary and secondary NB tumours mainly from Swedish patients. In addition, we have used nine NB cell lines and control tissues from blood lymphocytes and normal adrenal tissue.

The tumour cell content of the samples was histologically assessed in tumour tissue adjacent to that used for DNA extraction. Genomic DNA was extracted from NB cell lines or from fresh NB tumour tissue, fresh frozen (-70 °C) or formalin fixed paraffin embedded (FFPE) tissue with a DNeasy blood and tissue kit (Qiagen, Hilden, Germany), according to the protocol provided by the supplier, or was phenol extracted using phase lock gel (Eppendorf AG, Hamburg, Germany).

### *Paper I*

- Tumour DNA from 54 paired diagnostic– relapse NB patients

### *Paper II*

- Tumour DNA from 356 NB patients
- Cell lysate from eight NB cell lines

### *Paper III*

- Tumour DNA from 105 NB patients were used

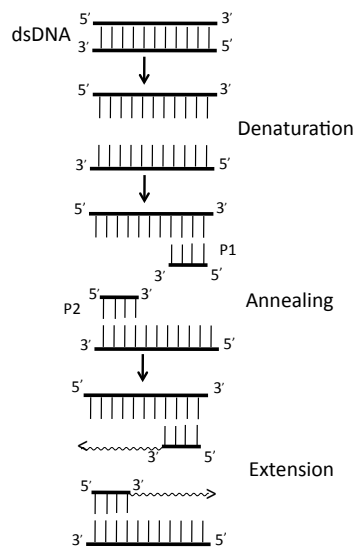
### *Paper IV*

- Tumour DNA from 535 NB patients were used
- Cell lysate from seven NB cell lines (SK-N-BE, Kelly, IMR-32, NB1, CLB-BAR, CLB-GE, CLB-GA)

# METHODS

## Polymerase Chain Reaction (PCR)

PCR has revolutionized molecular genetics since its introduction in the mid-1980s (Mullis et al 1986). PCR is used for multiplying shorter stretches of a defined target DNA sequence within a heterogeneous collection of DNA sequences. A PCR run consists of a series of cycles comprising three steps, (I) the denaturation of the dsDNA template, (II) annealing of primers and (III) DNA synthesis (extension), see Figure 9. DNA is elongated from the bound primers in the presence of heat-stable DNA polymerase and deoxynucleoside triphosphates (dNTPs). The newly synthesized DNA strands will act as templates in following cycles. A PCR run can be divided into three phases. During the first phase, the exponential phase, the PCR product is doubled at every cycle. The next phase is the linear phase in which the reaction slows down because the reaction components are consumed and the product starts to degrade. In the last phase, the plateau phase, the reaction has stopped and no more products are made. In traditional PCR, the products of the final phase are often evaluated on agarose gel after staining with Gel Red to estimate size, specificity and concentration. This is often called end point detection. “Hot start” Taq DNA polymerase for the PCR amplification of DNA has mainly been used in this thesis. These types of Taq polymerase utilize different techniques to keep the polymerase inactivated until the initial heat activation step, which results in more specific, selective PCR amplification.

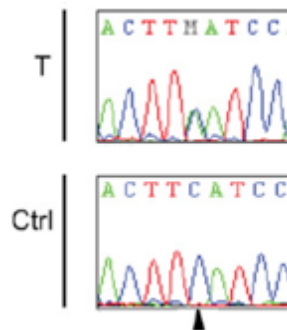


**Figure 9.** Schematic drawing of PCR amplification

## Mutation screening methods

### *Sanger sequencing*

DNA sequencing is the determination of the precise sequence of nucleotides in a template of DNA. Historically, the most common method of sequencing, originally developed by Sanger et al in 1977, is called the Sanger method or dideoxy method. In this method, fluorescently labelled nucleotides that lack the OH at the 3' carbon atom (dideoxynucleotides, ddNTP) are added to a purified PCR product, together with normal deoxynucleotides, DNA polymerase I and one primer. The ddNTPs are randomly incorporated and terminates the elongation chain because of the lack of 3' OH for the next nucleotide to be attached to. For this reason, the dideoxy method is also called the chain termination method. The sequence reaction will result in nucleotide chains of different length, from chains ending directly after the forward primer to chains covering the whole amplicon. The nucleotide chains from the sequencing reaction are then precipitated and resuspended in a denaturing formamide solution to keep them single stranded and separated on an automate sequencer where laser excitation of the terminal fluorescent dyes make it possible for a software program to present the bases as peaks in an electropherogram. In this thesis, Big Dye Terminator ddNTP chemistry from Applied Biosystems was used to sequence PCR products from genomic DNA. PCR purification and sequencing precipitation were conducted with AMPure magnetic beads and CleanSeq magnetic beads, both Agencourt (Agencourt Bioscience Corporation), using the Biomek NX pipeting robot (Beckman Coulter). Sequencing was performed on 3730 (Life Technologies) in house or at GATC biotech AG, European Custom Sequencing Center (Germany). Part of sequencing in paper I to III was performed with Sanger.



**Figure 10.** Example of DNA sequencing electropherogram. Mutation (F1245L) in the *ALK* gene in a NB tumour, upper sequence and reference DNA, bottom sequence.



### *Principles of Massively Parallel Sequencing (MPS)*

Sequencing has been going through a revolution in the last years. Since 2005, massively parallel sequencing (MPS), alias next-generation sequencing (NGS), is making its way from research laboratories into applied sciences and clinics. This technology uses miniaturized and parallelized platforms for sequencing of 1 million to 43 billion short readouts (50-400 bases each) per instrument run (Moorthie et al 2011). Many NGS platforms differ in engineering configurations and sequencing chemistry. They share the technical paradigm of massive parallel sequencing via spatially separated, clonally amplified DNA templates or single DNA molecules on a solid surface either on beads or a flat glass of micro fluidic channels. This design is very different from that of Sanger sequencing that is based on electrophoretic separation of chain-termination products produced in individual sequencing reactions. Fundamental to NGS library construction is the preparation of the nucleic acid target, RNA or DNA, into a form that is compatible with the sequencing system to be used. For targeted re-sequencing there are mainly two technical preparation methods to construct a library of fragments. The first is Amplicon-based enrichment assay involving a multiplex PCR reaction with multiple primer pairs. Primers include all sequences needed for downstream analysis. The second approach is a two-step process, first a sample preparation step of the genomic DNA and then a fishing step where the genomic regions of interest are captured by hybridization-based technique. The library construction on the second approach starts with DNA fragmentation that can be done either by restriction enzyme digest, by using transposase or by the Covaris, which is a focused ultrasonicator. Then synthetic DNA adapters are covalently added to each fragment end by DNA ligase. The adapters in general contain three parts: universal sequences specific to each platform, sequence used to amplify the DNA and indexes (also called tags or barcodes) used for distinguishing samples when pooling several samples together. In MPS, sequencing and detection is done simultaneously and include three steps: addition of a nucleotide, detection of the nucleotide incorporate and finally a washing step that may include chemistry to remove fluorescent labels or blocking groups.

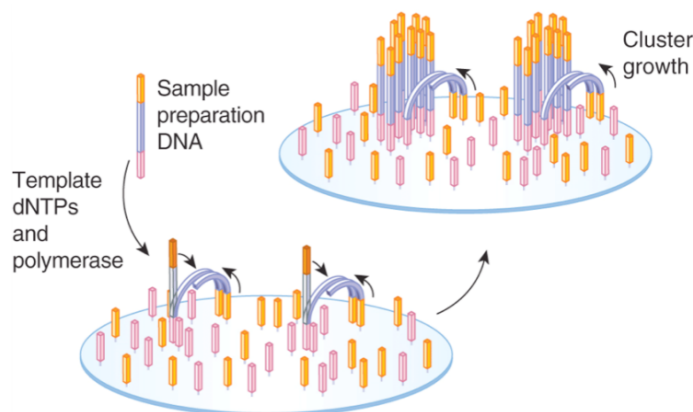
### *Bridge amplification and sequencing by synthesis (Illumina)*

This sequencing system was initially developed by Solexa in 2007 and subsequently acquired by Illumina. The Illumina flow cell is composed of flat glass with covalently attached adapter sequences complementary to the library adapters. A diluted solution of the library fragments are immobilized and amplified in situ on the flow cell surface by using bridge amplification. The amplification is digital, meaning that each DNA molecule is amplified in a cluster that is delimited from other DNA molecules. The solid-phase amplification creates up to 1,000 identical copies of each single template molecule in close proximity: a “cluster”, whose diameter is 1 micron or less. This amplification is necessary to provide sufficient signal from each fragment.

The fragmented genomic DNA has ligated adapters to both ends of the fragments. These fragments will be single-stranded and hybridize randomly to the dense lawn of primers existing on the inside surface of the flow cell channels. These primers are strand complementary to the adaptor sequences (T5 & T7) and by adding unlabelled nucleotides and enzyme the extension initiates. The double stranded molecule is denatured and the original template is washed away. Denaturation leaves single-stranded templates anchored to the substrate. Then the single strand folds over and the adapter region hybridizes to second type of primer on the flow cell. The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate. This bridge is denatured resulting in two single stranded copies of the molecule that are tethered to the flow cell. The process is then repeated over and over and occurs simultaneously for a millions of cluster resulting in clonal amplification of all fragments. After bridge amplification the reverse strands are cleaved and washed off leaving only the forward strands. The 3' ends are blocked to prevent unwanted priming. Sequencing begins with the extension of the first sequencing primer to produce the first read. Within each cycle four labeled reversible terminators nucleotides compete for addition to the growing chain, only one is incorporated based on the sequence of the template. After the addition of each nucleotide the clusters are excited by a light source and a characteristic fluorescent signal is emitted and captured from each cluster. This process is called sequencing by synthesis. The number of the cycles determines the length of the reads. The emission wavelengths along with the signal intensity determine the base call. For a given cluster all identical strands are read simultaneously. Hundreds of millions clusters are sequenced in a massively parallel process. After the completion of the first read, the read product is washed away. In this step, the index 1 primer is introduced and hybridizes to the template. The read is generated similar to the first read. After the completion of the index read, the read product is washed off and the 3' end of the template is deprotected. The template now bends over and binds the second primer on the flow cell. Index 2 is read in the same manner as index one. Index 2 read product is washed off at the completion of the step. Polymerase extends the second flow cells primer forming a double stranded bridge. This double stranded DNA is then linearized and the 3' end is blocked. The original forward strand is cleaved off and washes away leaving the reverse strand. Read two begins with the introduction of read 2 sequencing primer. As with read one the sequencing steps are repeated until the desired read length is achieved. The read-two product is washed away. This entire process generates millions of reads representing all the fragments. Sequences from pooled sample library are separated based on the unique indexes introduced during sample preparation.

For each sample, reads with similar stretches of base-calls are locally clustered. Forward and reverse reads are paired creating contiguous sequences that will be aligned back to the references genome for variant identification. The pair-end information is used to resolve ambiguous alignment. However, Illumina has its own systematic base calling biases. Most importantly, different tiles of the sequencing plate tend to produce reads of different quality (Dolan & Denver 2008); the 3' ends of sequences tend to have higher sequencing error rates compared to the 5' ends

(Schroder et al 2010), and increased single-base errors have been observed in association with GGC motifs (Nakamura et al 2011). Another source of noise is dephasing where the different copies of a DNA fragment within a cluster get out of synch. If the deblocking doesn't occur on a particular strand, then that strand won't incorporate the next nucleotide. But it might be deblocked in the next round, in which case it will be one base behind the rest of the cluster. At this point, the imaging of the cluster will result in a mix of bases, so the quality will be low for the basecall. The Illumina platform was used for amplicon-based sequencing in paper III.



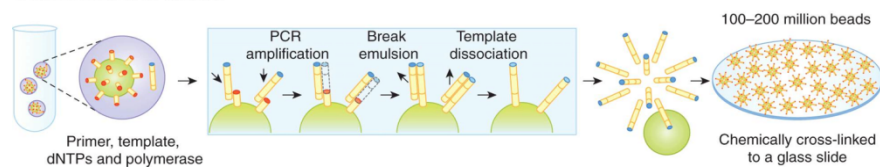
**Figure 11.** Bridge amplification. Enhance signalling through increasing the copy numbers of a single fragment. From (Metzker 2010).

### *Ion Torrent Personal Genome Machine (PGM™) technology*

Life Technologies expanded their NGS portfolio with the acquisition of Ion Torrent and their first system, the Ion Personal Genome Machine (PGM) sequencer, in 2010.

During library construction according to Ion Xpress Fragment Library Kit, PCR products of relevant genomic regions are end-repaired and Ion-Torrent barcoded adapters are ligated at the 5' and 3' ends. The fragments generated during the library prep are attached to beads and amplified using emulsion PCR (emPCR). Beads coated with complementary primers are mixed with a dilute aqueous solution containing the fragments to be sequenced along with the necessary PCR reagents. This solution is then mixed with oil to form an emulsion of microdroplets. The concentration of beads and fragments is kept low enough such that each microdroplet contains only one of each (or possibly none, but almost never more than one). Clonal amplification of each fragment is then performed within the microdroplets. The beads are put into a picotiter plate that holds a single bead in

each of several hundred thousand single wells married to ion sensor, which provide a fixed location at which each sequencing reaction can be monitored. Following amplification the emulsion is ‘broken’ (generally by organic extraction and centrifugation) and the amplified beads are enriched in a glycerol gradient, with unamplified beads pelleting at the bottom. What really differentiates Ion Torrent’s systems is the sequencing technology. It is based on the standard pyrosequencing chemistry, a form of ‘sequencing by synthesis’ whereby individual bases are introduced one at a time and incorporated by DNA polymerase. However, unlike other platforms based on pyrosequencing, rather than measuring light released from chemiluminescent reagents, the Ion Torrent system measures the direct release of H<sup>+</sup> (hydrogen ion) from the reaction. As the sequencer floods the chip with one nucleotide after another, any nucleotide added to a DNA template will be detected as a voltage change, and the PGM™ System will call the base. Because optics isn’t required, they’re able to make relatively inexpensive instruments coupled with disposable chips, which essentially act as pH meters. The lack of optics also means they don’t have to contend with slow image scans, so the sequencing reactions are relatively fast, with 200b reads taking about 2 hours. Finally, the lack of fluorescence or chemiluminescence means that the system can use unmodified nucleotides, which are cheaper and better tolerated by DNA polymerase. While the error rates are generally pretty good (~1%), pyrosequencing chemistry has trouble with long homopolymers (stretches of the same base, e.g. AAAAAA). Because the chemistry does not pause after each base incorporation, stretches of the same base will result in a single, albeit stronger, signal. While short stretches can be differentiated, it becomes increasingly difficult with longer stretches. The PGM technology has been used in paper I.

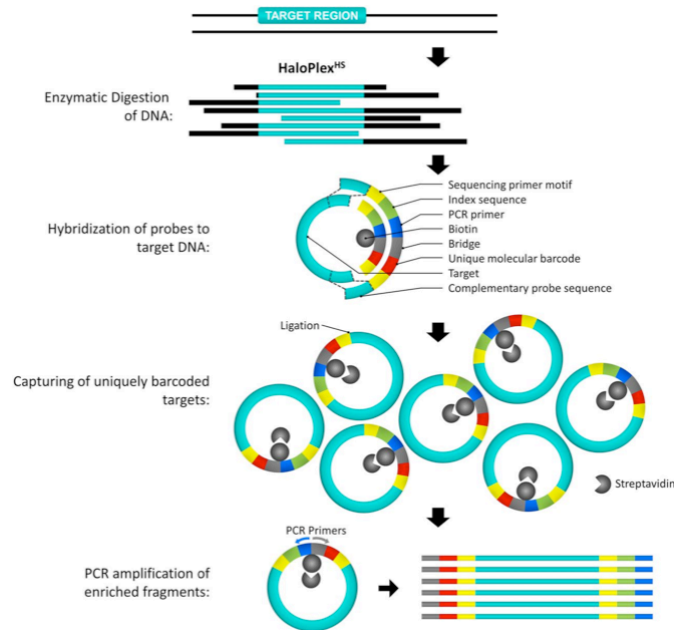


**Figure 12.** *Emulsion PCR. From (Metzker 2010).*

### *HaloPlex™ Target Enrichment System*

Agilent acquired HaloPlex platform in December 2011. HaloPlex™ Target Enrichment System enables simple and efficient analysis of genomic regions of interest in a large number of samples. Custom designs, for panels containing thousands of exons, are easily created with SureDesign. Using a simple workflow, a DNA sample is fragmented using restriction enzymes, and denatured. The probe library is added and hybridized to the targeted fragments; probes can be designed to any target sequences. Each probe is an oligonucleotide designed to hybridize to

both ends of a targeted DNA restriction fragment, thereby guiding the targeted fragments to form circular DNA molecules. The ends of the probe are complementary to the desired fragment sequence and the middle part of the probe is composed of a sequencing specific motif that is incorporated during circularization. In addition, a sample barcode sequence is incorporated in this step. The HaloPlex probes are biotinylated and the targeted fragments can therefore be retrieved with magnetic streptavidin beads. The circular molecules are then closed by ligation, a very precise reaction that ensures that only perfectly hybridized fragments are amplified. Only circular DNA targets are amplified, providing an enriched and barcoded amplification product that is ready for sequencing. Part of sequencing reactions in paper III is performed with HaloPlex technique.



**Figure 13.** Schematic overview of library preparation and capture process according to the HaloPlex technology. From (de Kock et al 2016).

### Data processing

The term coverage (depth) is the average number of times that each nucleotide is expected to be sequenced assuming that reads are randomly distributed across the target region. The coverage can be calculated by  $\text{number of reads} \times \text{read length (output)} / \text{target size} \times \text{number of samples included}$ . Uniformity of coverage will be influenced by low-complexity sequences, which is characterized by a significant

proportion of reads sharing identical start sites. This result in a lot of redundant sequence reads which just end up in the trash.

Multiple algorithmic approaches have been developed for analysing the output data from NGS but it includes the main steps: base calling, quality control, alignment and variants calling. Sequencing platforms typically have integrated base calling software. Base quality score, a measurement of uncertainty to each base call from image analysis, is typically given in the standard Phred quality score (Phred score 20 corresponds to 1% error rate in base calling). Checking the quality of generated read data is often the first step in a pipeline. Read trimming is often done by removing poor quality reads, particularly near the sequencing primer site, and toward the end of longer sequence runs. BWA (Burrow-Wheeler transform algorithm) is short read aligner that has been used in this thesis for alignment of reads to a reference genome, human genome build hg19. A good alignment is important for the next step of variation detection. Sequence alignments have to produce well-calibrated mapping quality scores to predict the accuracy of an observed variant. Sequence alignment map (SAM) format data is output from aligners that read FASTQ files and assign the sequences to a position with respect to a known reference genome. The binary version of SAM is called BAM and the SAM/BAM format, has become the standard format for storing the result of alignment step that can be used by downstream tool kits. Visualization of the alignment can be done in the Integrative Genomics Viewer (IGV). It is recommended to remove or at least mark PCR duplicates, arising during library amplification, before proceeding to variant calling step.

A variant call is a conclusion that there is a nucleotide difference versus reference at a given position in an individual genome. It is usually accompanied by an estimate of variant frequency and measure of confidence. Genotyping identifies the set of alleles present at each locus and this is only done for positions where variants already have been called. The accuracy of a variant calling depends on mapping quality, read depth and allele balance. Ploidy should be considered for cancer genome samples. GATK (Genome Analyzer Tool Kit) is the current best probabilistic method producing robust estimates of the probabilities of each of the possible genotypes. Variant Call Format (VCF) is a standardized text file format for representing SNP, indel, and structural variation calls. Furthermore, genome annotation is the process of identifying the locations of genes, coding regions and other sites of interest in a DNA with descriptive information about function. Once a genome is sequenced, it needs to be annotated to make sense of it.

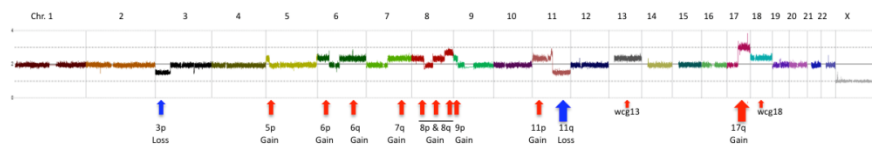
### *Single Nucleotide Polymorphism (SNP) array*

Single nucleotide polymorphisms (SNPs), the most common variation at a single site in the human genome, refer to genomic positions where two or more bases are found in the population. With an estimated frequency of about 10 million SNPs, evenly dispersed across the human genome, they represent a suitable target for large-scale association studies aiming to find the causing genes for the disease in question (International HapMap Project, 2003). The assembly of a large number of SNP assay to an array was first described in 2003 (Kennedy et al 2003). Although this technique was first used for the purpose of high throughput genotyping, it was soon recognized that these arrays could also be used for copy number analysis (Bignell et al 2004; Zhao et al 2004).

In high-density oligonucleotide SNP arrays, hundreds of thousands of probes are arrayed on a small chip, allowing for many SNPs to be interrogated simultaneously. SNP arrays use short oligonucleotides (~25nt) that are located such that a particular SNP is covered (Speicher & Carter 2005). Because SNP alleles only differ in one nucleotide and because it is difficult to achieve optimal hybridization conditions for all probes on the array, the target DNA has the potential to hybridize to mismatched probes. This is addressed somewhat by using several redundant probes to interrogate each SNP. Probes are designed to have the SNP site in several different locations as well as containing mismatches to the SNP allele. By comparing the differential amount of hybridization of the target DNA to each of these redundant probes, it is possible to determine specific homozygous and heterozygous alleles. Separate probes are synthesized to match each of the possible alleles, this enabling genotyping of the SNP by comparing the fluorescent intensity between the two sets of probes. In addition to genotyping, the fluorescent intensity for each probe also enables the copy number to be inferred from the array (Bignell et al 2004; Zhao et al 2004). For SNP arrays a separate intensity measurement is present for each of the alleles, thus enabling allele specific copy number to be analysed in addition to the total copy number. This feature makes it possible to detect both loss of heterozygosity (LOH) and copy number changes simultaneously (Sato-Otsubo et al 2012; Speicher & Carter 2005). SNP arrays are performed as single channel experiments, with only a single sample hybridized to each array. The fluorescent intensities are then compared *in silico*, either to a set of reference samples from healthy individuals or to the matched control sample from the same patient. The resulting copy number plot then shows the change in copy number relative to the controls used.

In this thesis, the tumours were analysed for copy number changes and LOH using either Genechip® Human Mapping 50K, 500K NspI, SNP array 6.0 or CytoScan® HD (Affymetrix Inc., Santa Clara, CA). Here the most used array is the Mapping 500K from Affymetrix that is comprised of two arrays, each capable of genotyping on average 250,000 SNPs. The array that uses the NSP I restriction

enzyme (approx. 262,000 SNPs) has been used here, with an average distance between SNPs of 11 kb. We also used CytoScan® HD, 750,000 SNP probes and 1.9 million non-polymorphic probes, with a mean marker distance of 1 SNP per Kbp that has highest resolution to accurately detect breakpoints, LOH, regions of identical-by-descent, maternal contamination, and low-level mosaicism. The array experiments were performed at our lab according to the protocol provided by the supplier (Affymetrix, Inc., Santa Clara, CA). Briefly, total genomic DNA (250 ng) was digested with the restriction enzyme and ligated to adaptors. After ligation, the template was subjected to PCR amplification using a generic primer that recognizes the adaptor sequence. PCR conditions have been optimized to preferentially amplify fragments in the 200 to 1,100 bp size range. The purified PCR products fragmented with DNase I, end labeled with biotin by using the enzyme Terminal Deoxynucleotidyl Transferase (TdT) before hybridized to a GeneChip Human Mapping array. The hybridized probes were washed using the Affymetrix Fluidics Station 450 and marked with streptavidin-phycoerythrin. The arrays were scanned using a confocal laser scanner, GeneChip Scanner 3000 (Affymetrix, Inc., Santa Clara, CA). To ensure proper detection of LOH regions in the case of tumour heterogeneity or containing significant normal cell components, we have used the AsCNAR algorithm available in CNAG 3.0, 3.3 and 3.5.1. When calculating the allele specific intensities, this algorithm takes advantage of the false heterozygous calls that might appear as a result of contaminating normal cells. This makes it a reliable method for genome wide detection of LOH and other allelic imbalances even in complex tumour materials (Nannya et al 2005; Yamamoto et al 2007). The ChAS software 2.1 and 3.1 (Affymetrix) was also used in the analyses.



**Figure 14.** Results of a NB tumour analyzed with the Affymetrix SNP array, visualized with GCOS software. The patient displays 11q deletion genomic profile with several segmental aberrations. Gains (red arrows) and losses (blue arrows). wg, whole chromosome gain

#### *Data Analysis & software*

GeneChip® operating software (GCOS) and GeneChip® Genotyping Analysis Software (GTTYPE, Affymetrix) were used for primary data analysis, normalization against internal control features on the chip, genotype calling and quality control. Subsequent analysis was then performed using Copy Number Analyzer for GeneChip (CNAG 3.0, Genome Laboratory, Tokyo University) featuring the algorithm for Allele-specific Copy-Number analysis using Anonymous References



(AsCNAR) and also ChAS software 2.2 (Affymetrix) was also used. In CNAG, the tumour samples were compared in silico to the 6 best-matched control samples (lowest standard deviation) available among a set of non-matched healthy individuals. This set contained both HapMap samples available from Affymetrix as well as our own set of healthy control samples. All cases of chromosomal gain, loss, or amplification were scored for both segmental and numerical aberrations, including detailed information about the breakpoint positions when applicable. The UCSC genome browser, assembly February 2009 <http://genome.ucsc.edu/>, was used to visualize gene regions.

## Cell culture and immunoblotting

Neuroblastoma cell lines were cultured in RPMI-1640 or DMEM medium supplemented with 10% foetal bovine serum (FBS) and 1% penicillin and streptomycin at 37°C, 5% CO<sub>2</sub>, 95% humidity. Cells were harvested in lysis buffer (25 mM of Tris-Cl, pH 7.5, 150 mM of NaCl, 1% (v/v) Triton X-100, 1 mM DTT, protease inhibitor cocktail tablet (Roche)). Cell lysates were cleared by centrifugation at 14,000 rpm for 15 minutes at 4°C, after which the samples were boiled in SDS sample buffer, and subsequently were run on SDS-PAGE, followed by immunoblotting using ALK (D5F3, Cell Signalling Technology), ALKAL1 (HPA027368, Atlas Antibodies), DLG2 and Tubulin (#2144, Cell Signalling Technology) antibodies. Horseradish-peroxidase HRP conjugated goat anti-mouse IgG and goat antirabbit IgG secondary antibodies (1:5000) were from Thermo Scientific.

## Statistical methods

### *Fisher's exact test*

Fisher's exact test, which is a hypothesis test of small samples that is an alternative to the Chi squared test, is used in the analysis of 2x2 association tables. The method is used in paper I to compare percentages of base frequencies (allele fractions) between data sets, i.e. for a given base between a case and the controls.

### *Kaplan-Meier survival analysis*

The probability of surviving a given length of time can be calculated by applying Kaplan-Meier method. The survival curve is a series of declining horizontal steps (e.g. dead of disease). On the plot, small vertical tick marks indicate individual patients whose survival times have been censored (e.g. alive with or without

disease). In paper III and IV survival probability rates were calculated using this approach.

# RESULTS AND DISCUSSION

## *Paper I*

Many NB patients experience relapse after successful initial treatment. Knowing more about the molecular properties of tumours at the time of relapse is a prerequisite to benefit from targeted molecular therapy. However, the actual frequency of *ALK* actionable mutations at relapsed neuroblastoma has not been studied at the time of this paper. To do this, we analysed *ALK* Sanger sequencing data for 54 paired neuroblastomas profiled either at diagnosis or at disease relapse matched with constitutional DNA. *ALK* mutations were observed in nine of 54 diagnostic samples. For all nine cases, the same mutation was also detected by Sanger sequencing in a sample obtained at the time of relapse. In one case, an *ALK* mutation, detected at diagnosis was present in at relapse in only one of several samples. Moreover, the *ALK* mutation was found in only one of several stroma-poor nodules having high mitosis-karyorrhexis index (MKI) and not in the other nodul or in the stroma-rich component. In 5 additional cases, de novo *ALK* mutations were also detected by Sanger only at the time of relapse. These relapse-specific mutations were a F1174L mutation (exon 23) in two cases, a F1174S mutation (exon 23) in one case, and a Y1278S mutation and R1275Q mutation (exon 25) in the other two cases.

Applying IonTorrent PGM technique on these cases, we then sought to determine whether the *ALK* mutation might in fact have gone undetected in the diagnosis samples due to a limit in sensitivity of the Sanger technique. We sequenced the relevant *ALK* hotspots in four of five cases with an *ALK* mutation present only at relapse. For the fifth case, no sufficient material was available. In 2 cases with an *ALK* mutation detected in the relapse sample, PGM analysis of diagnosed sample showed no evidence of mutated allele presence. Interestingly, in two other cases, PGM analysis revealed the base corresponding to a mutated allele in the diagnosis sample in a significantly higher frequency than the controls. In one of these two cases, an *ALK* mutated subclone was identified at diagnosis at the frequency of 0.79%. Strikingly, in another case an *ALK* mutation distinct from the one observed at relapse, but leading to the same AA change, was observed in 8.15% at diagnosis. We believe that a mutational switch with an expected identical functional consequence has happened relating this case.

Finally, to search for further evidence of clonal selection of *ALK* mutant cells, two established cell lines (CLB-Ma and CLB-Ba), with a previously described F1174L mutation observed by Sanger sequencing were compared to the corresponding primary sample. For CLB-Ma PGM™ analysis of the primary tumour tissue did not reveal any *ALK*-mutated subclone. However, the cell line has been established from

the infiltrated bone marrow. Since there was no sample from this bone marrow available for PGM analysis, the presence of *ALK*-mutated subclone in a metastatic site cannot be excluded. For CLB-Ba, on the other hand, the PGM analysis of invaded bone marrow from which the cell line was directly established showed the presence *ALK* mutation in 6.6% of all reads. This is an evidence for the presence of an *ALK*-mutated subclone in the primary sample which expanded during establishment of the cell line contributing to a selective advantage. We have shown that the *ALK* mutational events increase at relapse, likely as a result of clonal evolution of mutation-carrying cells during primary treatment. The potential of targeted molecular inhibitors justifies repeated tumour sampling as a new standard of care.

Our data suggest that neuroblastomas undergo substantial mutational evolution during therapy, and that relapsed disease is more likely to be driven by a targetable oncogenic pathway, highlighting that it is critical to base treatment decisions on the molecular profile of the tumour at the time of treatment. However, it will be necessary to conduct prospective clinical trials that match next-generation sequencing results to targeted therapeutic intervention to determine if cancer genomic profiling improves patient outcomes.

## *Paper II*

Segmental chromosome alterations (SCAs) most often involving chromosome regions 1p, 1q, 2p, 3p, 4p, 11q and 17q are frequently observed in advanced stages of NB. However, frequently detected gene alterations are limited to *MYCN* amplification (20%) and *ALK* activations (7%). *ALK* interplay with *MYCN* in NB pathogenesis as *ALK* mutation or amplification, most commonly affecting one of the three hotspot residues in the *ALK* kinase domain (F1174, R1245 and R1275), happens predominantly in *MYCN*-amplified NB cases.

The human *ALK* locus, spans 728kb (chr2:29,192,774-29,921,566), as well as *MYCN* locus, spans 4,4kb (chr2:15,940,564-15,947,007), are found at the distal portion of 2p, where it is frequently the target of amplification in NB (GRCh38/hg38 Assembly). High-level amplifications of *ALK* co-occurs with *MYCN* amplification in an estimated 2% of NB while focal amplification of *ALK* alone is a rare event (Caren et al 2008a; Mosse et al 2008). Furthermore, the recently described *ALK* ligand *ALKAL2*, spanning 8,7kb in chr2: chr2:279,558-288,296, can also be found within this 30Mb distal portion of chromosome 2p, along with *ALK* and *MYCN* (Guan et al 2015; Reshetnyak et al 2015). What advantages does 2p gain provide in the development of NB? An obvious first candidate is *MYCN* oncogene representing one of the most powerful prognostic markers in NB malignancy (Ambros et al 2009). The tumourigenic potential of MNA process is most probably a consequence of an increase in gene dose in a range of 40-200 copies which is a highly demanding process for the cell (Levan et al 1976). An over fourfold increase in the *MYCN* signals number relative to chromosome 2q indicates

a *MYCN*-amplified tumour (Ambros et al 2003). In order for *MYCN* amplification that increases gene expression to contribute to adaptive evolution, the fitness advantages of increased gene expression must outweigh any decrease in fitness associated with the energetic costs of increased gene expression. This fact argues against gain of one single copy – i.e. going from two to three copies, as in the case of 2p gain, being sufficient for promoting tumour development/progression.

The recent discovery of the *ALK* ligands (Guan et al 2015; Reshetnyak et al 2015), one being *ALKAL2* (FAM150B/AUG $\alpha$ ), with a genetic locus in the genomic proximity of *MYCN* and *ALK*, makes it tempting to speculate that the 2p gain in NB tumours is not solely due to the tumour benefitting from an extra copy of *MYCN* but instead from an extra copy of a NB promoting 2p gain cassette including three genes i.e. *ALKAL2*, *MYCN* and *ALK*.

In this paper, we analysed a large set of NB tumours (n=356) for presence of segmental aberration including amplification and segmental loss or gain. The extent of aberrations on 2p was defined and for clarity, we also included *ALK* mutation data where possible. Overall 107 of the 356 NB cases examined exhibited either gain or amplification of 2p. The most common gain on 2p contains all three genes which lends support for the hypothesis that gain of *ALKAL2*, *MYCN* and *ALK* provides an advantage to NB growth. In addition, we observed that *ALK* mutation is almost exclusively present in the *MYCN*-amplified group.

Since it was of interest to analyse for mutations in the *ALK* gene in the group of NB tumours with rearrangements in 2p we subjected available samples to Sanger sequencing for *ALK* tyrosine kinase domain covering exons 21-25. 8 out of 107 cases showed *ALK* amplification. 67 out of 107 patients were negative for *ALK* mutation; 11 out of 107 cases (10.2 %) were positive for *ALK* mutation detected by Sanger method; They were all in either of positions 1174, 1245 or 1275, which are the most common positions for *ALK* mutation in NB tumours. *ALK* sequencing was not performed in 21 patients due to lack of material. None of the cases with *ALK* amplification carried an *ALK* mutation. High levels of *ALK* expression, due to mutation or amplification, can lead to ligand independent constitutive oncogenic activation in cancer cells; however, such findings suggest that ligand dependent activation of the wildtype *ALK* receptor concomitant with *MYCN* protein stabilization may play an important role in NB pathogenesis. The question of whether therapeutic options targeting *ALK* signalling should be considered in 2p gain NB remains to be addressed.

### *Paper III*

In the following study, we wanted to expand our investigation of the potential occurrence of *ALK* tyrosine kinase domain hotspot mutations that might be present at an allele frequency below the detection limit of Sanger sequencing technique. Tyrosine kinase inhibitors have become the gold standard therapy in

treating *ALK* positive NBs, making diagnostic high sensitive detection of *ALK* point mutations a necessary step in identifying optimal treatment modalities. In a clinically representative series of 105 NB samples, point mutational status of the *ALK* gene within exons 23, 24 and 25 (exons 21-25 in HaloPlex assay) containing the F1174, R1245 and R1275 mutation hotspots respectively, were sequenced using an amplicon-based deep sequencing assay (5000-30000X).

*ALK* point mutations were detected in sixteen out of the 105 (15.2%) tumour samples while gene amplification was noticed in additional four samples, giving a total of *ALK* mutational events in 20 of 105 neuroblastoma samples (19%).

At the I1171 locus (chr2:29445213), an *ALK* variant was detected in one case with the frequency of mutated allele as low as 2.7%. This was the lowest variant frequency level detected by deep-targeted sequencing in our cohort. At the F1174 hotspot (chr2:29443695-29443697), alterations were observed in eleven cases: seven cases harboured a mutation leading to the amino acid change F1174L, two cases with F1174I, one each of the F1174C, and F1174S substitutions, with the MAF ranging from 14% to 60%. Interestingly the patient harbouring F1174S mutation with 58% frequency, in parallel contained a small subclone of F1174I mutation with 8% frequency of the mutated allele. At the F1245 hotspot (chr2:29436858-29436860), alterations were detected in 3 tumours: two samples showed a F1245I mutation while the third case carried the F1245C mutation, with frequencies of 14%, 51% and 52%, respectively. At the L1240 locus (chr2:29436875) a L1240V substitution was detected in one patient with a MAF at 57%. Altogether *ALK* mutations with a variant allele frequency less than 20% were detected in six tumours while ten other cases showed *ALK* mutations with higher variant allele frequency (between 20-60%). No additional variants outside previously reported mutational sites were observed in the targeted regions in our cohort.

Our results fail to reveal a distinct mutational spectrum in relation to NB different genomic subtypes (numerical only, other segmental, 17q gained or *MYCN*-amplified). However, no *ALK* mutations were detected in any of the tumours carrying segmental 11q-deletion (n=27). No statistically difference in overall survival was seen when comparing *ALK*-positive (including *ALK* amplified) neuroblastoma with neuroblastoma with wildtype *ALK*.

All four cases with *ALK* gene amplification also had co-amplification of *MYCN* and no *ALK* mutation was detected in any of these samples. Focal low copy number gain for *ALK* was also detected in one *MYCN*-amplified patient who also had a chromothripsis pattern for chromosome 2. The partial 2p gains encompassing the *ALK* locus were present in only 4 of 16 mutated tumours indicating that 2p gain is not a common mechanism for increasing mutated *ALK* copy number. This is while 2p gains are present in as much as 19.3% of all neuroblastomas with segmental imbalances (Caren et al 2008a). Interestingly, all patients (n=4) with partial 2p gain that also carries an *ALK* mutation is dead of disease. We observed that the majority of *MYCN*-amplified tumours (4 of 6) contained subclonal (<20% variant allele fraction) *ALK* aberrations while in

*MYCN*-nonamplified tumours the majority of *ALK* mutations (8 of 10) were clonal. This implies that *MYCN*-nonamplified tumours are dependent on the presence of this constitutive active kinase in a higher percentage of the cell population.

All *ALK* mutated samples were also tested by Sanger sequencing which confirmed all *ALK* mutations occurring with a mutated allele fraction above 15% leaving four samples with a mutated allele fraction below 15% undetected by Sanger sequencing in our hands. Three of neuroblastoma associated mutations, F1174L/I/S, I1171T, and F1245C, observed in our dataset are known to confer strong gain of function effect to *ALK* in vivo assays, whereas the F1174C, F1245I and L1240V function as oncogenic drivers in the progression of NB tumours are not yet characterized (Hallberg & Palmer 2013).

The post-treatment follow-up examination should include the *ALK* mutational status, since the elevated frequency of mutated allele values herald the appearance of a recurrence (Schleiermacher et al 2014). During this study we have observed that prognosis value of *ALK* as a single factor in primary NB tumours can be misleading because of the extremely heterogeneous prognosis based on the neuroblastoma patient's age, stage, and biology. These findings are of clinical importance given the potential role of *ALK* mutations in clonal evolution and relapse. The efficacy of ALK-inhibitors on cells harbouring *ALK* mutations at subclonal level remains to be determined.

#### *Paper IV*

Deletion of the long arm of chromosome 11 is one of the most frequent events that occur during the development of aggressive neuroblastoma. Clinically, 11q deletion is associated with higher disease stage and decreased survival probability. Chromosome 11, although average in size, is one of the most gene-rich chromosomes, an average gene density of 11.6 genes per megabase, including 1,524 protein-coding genes. Despite extensive efforts to identify the recurrently involved genes, and to understand the molecular mechanisms of 11q deletion, definitive answers are still unclear. In this article, it is our intent to provide a deeper understanding of 11q deletion role in the development and/or progression of NB.

To do this, we performed a comprehensive genome-wide analysis of DNA copy number in 535 NB tumours using 250 k and/or Cytoscan HD gene chip arrays from Affymetrix. 120/535 patients (22%) had deletion of the long arm of chromosome 11. Tumours with 11q deletion contained significantly more chromosomes with breaks compared to *MYCN*-amplified tumours and significantly more chromosomal breaks overall. This subgroup contains many additional recurring large-scale chromosome imbalances including loss of 4p, 9p, 14q, and gain of 1q, 7q, 2p, and 11p. While 11q deletion is more frequently detected than MNA (35-45% vs. 20-25%, respectively), both alterations are almost mutually exclusive, only 26 cases (4,7%) had both *MYCN* amplification and 11q deletion.

Two regions of the chromosome 11q SRO of deletions were identified; It is possible that the elusive 11q tumour suppressor gene could actually be more than one, and located in two separate regions, one larger and more proximal (85,3–103 Mb), associated with non *MYCN*-amplified tumours, and one smaller and more distal (111–118/qter), associated with amplified tumours. Thus, we suggest that different sets of 11q-deleted genes are important for the biological behaviour of the *MYCN*-amplified and the non-*MYCN* amplified cases (see Table 4 for a summary).

Chromosome 11 rearrangements	Frequency	SRO
Entire chr 11 loss	18% (n=98/535)	
Partial 11q loss	22% (n=120)	for MNA 111.0 -118.0 for MNA-neg 85.3 -103
Interstitial 11q loss	1,8% (n=10)	84.8-87.0
Homozygous 11q losses	(n=0)	
Gain/amp chr 11	10% (n=53)	68.5-70.7
Constitutional 11q loss	(n=1)	81.5–102.9

**Table 4.** Summary of findings in the study. MNA, *MYCN* amplification; SRO, shortest region of overlap

We observed that the 11q deletion rearrangements most often found to interfere with *DLG2* normal gene function by directly disrupting the gene structure (chr11: 83,455,012-84,923,589) or the breakpoint can occur approximately 50 kb upstream of the intact loci. Moreover, one of the 11-deleted samples showed a small interstitial 11q deletion harbouring 16 genes. Among these genes there are 2 candidates, *DLG2* and *CREBZF*, which show a significant lower expression level in unfavourable neuroblastoma compared to favourable in R2 database. In all different subsets of NB patients, there was a very strong correlation between *DLG2* low expression and worse outcome of disease, indicating a possible role for *DLG2* as a tumour suppressor gene (R2: microarray analysis and visualization platform (<http://r2.amc.nl>)). According to this along with the evidence coming from the tumors copy number variations (CNVs) screening, *DLG2* turns out to be the most proximal appropriate driver gene as the locus position overlaps the shortest region of overlap (SRO) for the deletion and it harbours recurrent somatic structural variations within the gene body. Disruption of the *DLG2* gene can be a target of LOH observed in neuroblastomas, which have lost the long arm from one of its chromosome 11. However, we could not find any supplementary genetic lesion in addition to the primary copy number variant observed in *DLG2* reflecting gene dosage sensitivity mechanisms.

To further investigate *DLG2* as a tumour suppressor gene in neuroblastoma we investigated the expression of *DLG2* protein in a range of neuroblastoma cell lines: CLB-BAR, CLB-GA, Kelly, IMR32, NB1, SKN-BE and CLB-GE. *DLG* protein levels were undetectable in most neuroblastoma cell lines tested, with the exception



of the CLB-GE cell line. We could see no significant difference in the DLG2 expression level between 11q-deleted and 11q-normal NB cells.

To examine DLG further, we employed the *Drosophila* system, where Dlg has been shown to function as a growth suppressor. Examination of the larval neuromuscular junctions revealed a co-localization of Dlg protein with the *Drosophila* Alk receptor homologue. However, inhibition of Alk signalling did not affect Dlg expression or localization.

In 53 out of 120 11q-deleted tumours, we could localize a large genomic gain or amplification at 11q13 proximal to the 11q-deleted region while others had an intact copy number change through the chromosome 11 until the appearance of terminal deletion. The consensus region spans approximately 2 Mb (68.5-70.7Mb) and contains Cyclin D1 (*CCND1*) candidate gene that is important for transition through the G1 checkpoint of the cell cycle.

We found no substantial difference in *MYCN* copy number between samples of neuroblastoma obtained from primary and metastatic sites from the same patient. In contrast, there can be differences in 11q status between diagnosis vs. metastasis samples obtained from the same NB patient. Here we observed three *MYCN*-amplified cases with intact chromosome 11 at the time of diagnosis, which had evolved 11q deletion after treatment procedure. Therefore, 11q segmental chromosomal loss cannot be the primary event during NB development. This particular observation underscores the opinion that 11q alterations are probably late and mostly secondary events in tumour evolution.

# CONCLUSION

The observation of new *ALK* mutations at relapse suggests that the frequency of *ALK* mutations maybe higher at relapse than at diagnosis.

Subclones harbouring *ALK* mutations may contribute to tumour evolution and relapse. This has major clinical implications.

Intratumour heterogeneity of *ALK* may hinder therapeutic strategies that depend on results from single tumour biopsy sample or single time point.

NB tumours with 2p gain benefits from an extra copy of a NB promoting 2p gain cassette including three genes i.e. *ALKAL2*, *MYCN* and *ALK*.

There were no significant survival differences observed in tumours with or without *ALK* mutations or amplifications.

Prognosis value of *ALK* as a single factor in primary NB tumours can be misleading because of the extremely heterogeneous prognosis based on the neuroblastoma patient's age, stage, and biology.

In our hands, no *ALK* mutation was detected in any of NB tumours carrying segmental 11q-deletion.

11q-deleted tumours with *MYCN* amplification had generally smaller 11q deletions than 11q-deleted tumours without *MYCN* amplification and the deleted region located more distally in this group.

11q-deleted NB tumours present not only loss of a putative tumour suppressor gene, *DLG2*, but also frequent gain of an oncogenic driver, *CCND1*. These findings have a potential clinical implication when it comes to increasing the accuracy of grouping and subgrouping patients with NB.

# FUTURE PERSPECTIVE

Growing evidence indicates that cancers contain many subclonal mutations, defined as mutations that are present in a few cells. These subclones are derived from the founding clone, and are defined by the additional mutations they carry, which are not present in the bulk population. The full portrait of cancer cell diversity over different sites and on different time points reveals a staggering complex contexture. There is intratumour diversity on the genetic, epigenetic, transcriptome and protein levels. We need to assess the degree to which genotypic and phenotypic heterogeneity correspond within a tumour. One important question is if subclonal mutations really have an effect on tumour cell phenotype? It is yet unclear whether such changes are merely consequences of increased proliferation under stress or have an important role in tumour growth through clonal collaboration?

Intratumour diversity as general may be a clinically useful parameter for risk assessment. However, it is not yet clear if advances in tracking the genetics of individual cancer cells precisely will be able to introduce clinically actionable targets and reduce the incidence of invasive cancers. The efficacy of targeted therapy approaches against a subclonal driver mutation present in a subset of cells within a tumour must be further investigated.

Evolutionary studies of cancer by analysing tumour types with a moderate level of genetic complexity is essential in order to obtain clearer information on basic evolutionary patterns. This allows us to plot subclonal hierarchies and to follow whether the specific architecture is preserved during dissemination to metastatic sites and up to the point of treatment resistant. Moreover, analyses of recurrent patterns of coexisting subclones will allow us to form hypotheses for subclonal collaborations. However, studying evolution from single time point is hard! Recently, liquid biopsies based on circulating cell-free DNA (cfDNA) analysis are described as surrogate samples for molecular analysis. The non-invasive nature of cfDNA sampling enables sequential analysis for tumour genetic profiling which shed light on the spatial and temporal genomic heterogeneity of neuroblastoma. The remaining challenge is to transform the expected dynamics of gene frequency changes into a more mechanistic theory of evolutionary genomics.

# ACKNOWLEDGEMENT

I would like to convey my profound gratitude to Professor *Tommy Martinsson*, my supervisor, for giving me the opportunity to work with this thrilling project! Thank you for years of guidance, interesting inputs, countless support and setting everything in perspective. I admire you for your knowledge and your personality! I feel honoured to have been a part of your visions.

I like to thank professor *Ruth Palmer*, my assistant supervisor, who provided me with excellent advices in both scientific and linguistics way, as well as sharing her broad knowledge and fruitful collaborations.

Special thanks go to Professor *Per Kogner* from Karolinska University Hospital for kindly supplying us with valuable clinical data of patients. Your outgoing and collaborative personality is exceptional.

All the present and former members of the neuroblastoma (NB) group, particularly *Rose-Marie Sjöberg* who I owe great debts of gratitude to for always. It is impressive how your sole presence makes things work in the lab and in real life! You radiate love and good energy. I wonder what would everything have been without you?!

*Susanne Fransson* for critical review and constructive comments on the scientific contents of manuscripts and this thesis. You inspire me on the field of science, CrossFit training and fashion! Enjoying your company, all the conference gala dinners have been a succé and memorable.

*Anna Djos* for your friendship and for encouraging me to start talking Swedish. You have always been there for me to translate Swedish conversations when I had no clue at all!

I also wish to express my appreciation to *Helena Carén* for her contribution to this work, and *Hanna Kryb* for the support she provided when I came to the group.

*Angela Martínez-Monleon* for cheering me up with funny stories and late-night conversations about how to embrace a different culture, hot Andalucía weather vs. rainy Sweden and more. I wish in near future we improve on being language exchange partners, Spanish vs. Swedish. Good luck with your own journey.

*Susanne Reinsbach*, the latest member of NB clan, for discussing and sharing your personal experience on the process of dissertation, but also for being eager and willing to give useful comments.

The staff from the Genomics Core Facility:

*Jennie Finnman* thanks for introducing me to principal methodologies used in this work, from Sanger Sequencing to Massively Parallel Sequencing and for sharing your experience. *Elham Rekabdar* for being warm and helpful during the course of my life as a PhD student. You devotedly stayed with me during late evenings to manage to mix endless number of DNA samples together in equimolar concentrations! *Erik Lorentzen & Katarina Truvé* for taking care of sequencing files and statistical assistance. *Malin Östensson* for assisting us on visualization of our breakpoints data and many things else.

To my co-workers on 4<sup>th</sup> and 5<sup>th</sup> floors of Clinical Genetics department creating a friendly and comforting environment especially *Ewa-Lotta Karrstedt & Maria Ybr* for great hospitality and helping with array protocols whenever unpredicted issues happened. *Anna Roblin* for great time travelling to USA. Your serene character and helpful mindset is lovely.

*Saideh Rajaei* for being a friend and sharing with me all information I needed to know as a PhD freshman. Your friendly attitude is appreciable.

Professor *Afrouz Bebboudi* for being an enthusiastic teacher; observing your strength and vivid spirit has been motivating.

*Sven-Öjvind Swahn* for valuable comments on the layout of grants applications as well as graphical issues of the actual thesis.

*Staffan Nilsson* for creating an amusing atmosphere in the house and for statistical analysis of our data.

This thesis is dedicated to my parents. I would like to thank my father, *Kia*, for his endless support, encouragement and inspiration. To my mother, *Zhila*, for her love, devotion and hopeful attitude irrespective of circumstances; And also to *Nima* for not only being a brother but a wise friend! Thanks for always believing in me.

*Navid Atlasi* thanks a million times for lightening up my life. I love your beautiful heart. You mean everything to me.

Finally, I am very much indebted to all children with neuroblastoma and their families for incredible contribution to this work. Because of you, I will forever be excited about neuroblastoma research.

This work was supported by grants from the BioCARE cancer research program, Swedish Cancer Society, the Children's Cancer Foundation, the Nilsson-Ehle Foundation, the Assar Gabrielsson Foundation, and the Sahlgrenska University Hospital Foundation.

# LITERATURE CITED

- Abel F, Ejeskar K, Kogner P, Martinsson T. 1999. Gain of chromosome arm 17q is associated with unfavourable prognosis in neuroblastoma, but does not involve mutations in the somatostatin receptor 2(SSTR2) gene at 17q24. *Br J Cancer* 81:1402-9
- Ambros IM, Benard J, Boavida M, Bown N, Caron H, et al. 2003. Quality assessment of genetic markers used for therapy stratification. *J Clin Oncol* 21:2077-84
- Ambros PF, Ambros IM, Brodeur GM, Haber M, Khan J, et al. 2009. International consensus for neuroblastoma molecular diagnostics: report from the International Neuroblastoma Risk Group (INRG) Biology Committee. *Br J Cancer* 100:1471-82
- Attiyeh EF, London WB, Mosse YP, Wang Q, Winter C, et al. 2005. Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N Engl J Med* 353:2243-53
- Basit S, Ashraf Z, Lee K, Latif M. 2017. First macrocyclic 3rd-generation ALK inhibitor for treatment of ALK/ROS1 cancer: Clinical and designing strategy update of lorlatinib. *Eur J Med Chem* 134:348-56
- Bauer A, Savelyeva L, Claas A, Praml C, Berthold F, Schwab M. 2001. Smallest region of overlapping deletion in 1p36 in human neuroblastoma: a 1 Mbp cosmid and PAC contig. *Genes Chromosomes Cancer* 31:228-39
- Berry T, Luther W, Bhatnagar N, Jamin Y, Poon E, et al. 2012. The ALK(F1174L) mutation potentiates the oncogenic activity of MYCN in neuroblastoma. *Cancer Cell* 22:117-30
- Bignell GR, Huang J, Greshock J, Watt S, Butler A, et al. 2004. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res* 14:287-95
- Bosse KR, Maris JM. 2016. Advances in the translational genomics of neuroblastoma: From improving risk stratification and revealing novel biology to identifying actionable genomic alterations. *Cancer* 122:20-33
- Bourdeaut F, Trochet D, Janoueix-Lerosey I, Ribeiro A, Deville A, et al. 2005. Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. *Cancer Lett* 228:51-8
- Bresler SC, Weiser DA, Huwe PJ, Park JH, Krytska K, et al. 2014. ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma. *Cancer Cell* 26:682-94
- Breslow N, McCann B. 1971. Statistical estimation of prognosis for children with neuroblastoma. *Cancer Res* 31:2098-103
- Brodeur GM. 2003. Neuroblastoma: biological insights into a clinical enigma. *Nat Rev Cancer* 3:203-16
- Brodeur GM, Pritchard J, Berthold F, Carlsen NL, Castel V, et al. 1993. Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. *J Clin Oncol* 11:1466-77
- Brodeur GM, Seeger RC, Barrett A, Berthold F, Castleberry RP, et al. 1988. International criteria for diagnosis, staging, and response to treatment in patients with neuroblastoma. *J Clin Oncol* 6:1874-81
- Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. 1984. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 224:1121-4

- Brodeur GM, Sekhon G, Goldstein MN. 1977. Chromosomal aberrations in human neuroblastomas. *Cancer* 40:2256-63
- Caren H, Abel F, Kogner P, Martinsson T. 2008a. High incidence of DNA mutations and gene amplifications of the ALK gene in advanced sporadic neuroblastoma tumours. *Biochem J* 416:153-9
- Caren H, Erichsen J, Olsson L, Enerback C, Sjoberg RM, et al. 2008b. High-resolution array copy number analyses for detection of deletion, gain, amplification and copy-neutral LOH in primary neuroblastoma tumors: four cases of homozygous deletions of the CDKN2A gene. *BMC Genomics* 9:353
- Cazes A, Louis-Brennetot C, Mazot P, Dingli F, Lombard B, et al. 2013. Characterization of rearrangements involving the ALK gene reveals a novel truncated form associated with tumor aggressiveness in neuroblastoma. *Cancer Res* 73:195-204
- Chen Y, Takita J, Choi YL, Kato M, Ohira M, et al. 2008. Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* 455:971-4
- Christensen JG, Zou HY, Arango ME, Li Q, Lee JH, et al. 2007. Cyto-reductive antitumor activity of PF-2341066, a novel inhibitor of anaplastic lymphoma kinase and c-Met, in experimental models of anaplastic large-cell lymphoma. *Mol Cancer Ther* 6:3314-22
- Cohn SL, Pearson AD, London WB, Monclair T, Ambros PF, et al. 2009. The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. *J Clin Oncol* 27:289-97
- Crick FH. 1958. On protein synthesis. *Symp Soc Exp Biol* 12:138-63
- De Brouwer S, De Preter K, Kumps C, Zabrocki P, Porcu M, et al. 2010. Meta-analysis of neuroblastomas reveals a skewed ALK mutation spectrum in tumors with MYCN amplification. *Clin Cancer Res* 16:4353-62
- de Kock L, Wang YC, Revil T, Badescu D, Rivera B, et al. 2016. High-sensitivity sequencing reveals multi-organ somatic mosaicism causing DICER1 syndrome. *J Med Genet* 53:43-52
- de Smith AJ, Walters RG, Coin LJ, Steinfeld I, Yakhini Z, et al. 2008. Small deletion variants have stable breakpoints commonly associated with alu elements. *PLoS One* 3:e3104
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, et al. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376-80
- Dolan PC, Denver DR. 2008. TileQC: a system for tile-based quality control of Solexa data. *BMC Bioinformatics* 9:250
- Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, et al. 1996. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med* 2:561-6
- Egger G, Liang G, Aparicio A, Jones PA. 2004. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429:457-63
- Esteller M. 2002. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* 21:5427-40
- Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, et al. 2014. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet* 23:5866-78
- Feinberg AP, Vogelstein B. 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301:89-92
- Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, et al. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529:110-4
- Fransson S, Hansson M, Ruuth K, Djos A, Berbegall A, et al. 2015. Intragenic anaplastic lymphoma kinase (ALK) rearrangements: translocations as a novel mechanism of ALK activation in neuroblastoma tumors. *Genes Chromosomes Cancer* 54:99-109

- Gilbert F, Feder M, Balaban G, Brangman D, Lurie DK, et al. 1984. Human neuroblastomas and abnormalities of chromosomes 1 and 17. *Cancer Res* 44:5444-9
- Griffith JS, Mahler HR. 1969. DNA ticketing theory of memory. *Nature* 223:580-2
- Guan J, Tucker ER, Wan H, Chand D, Danielson LS, et al. 2016. The ALK inhibitor PF-06463922 is effective as a single agent in neuroblastoma driven by expression of ALK and MYCN. *Dis Model Mech* 9:941-52
- Guan J, Umapathy G, Yamazaki Y, Wolfstetter G, Mendoza P, et al. 2015. FAM150A and FAM150B are activating ligands for anaplastic lymphoma kinase. *Elife* 4:e09811
- Guo C, White PS, Weiss MJ, Hogarty MD, Thompson PM, et al. 1999. Allelic deletion at 11q23 is common in MYCN single copy neuroblastomas. *Oncogene* 18:4948-57
- Hallberg B, Palmer RH. 2013. Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nat Rev Cancer* 13:685-700
- Hallberg B, Palmer RH. 2016. The role of the ALK receptor in cancer biology. *Ann Oncol* 27 Suppl 3:iii4-iii15
- Hallstenson K, Thulin S, Aburatani H, Hippo Y, Martinsson T. 1997. Representational difference analysis and loss of heterozygosity studies detect 3p deletions in neuroblastoma. *Eur J Cancer* 33:1966-70
- Hasan MK, Nafady A, Takatori A, Kishida S, Ohira M, et al. 2013. ALK is a MYCN target gene and regulates cell migration and invasion in neuroblastoma. *Sci Rep* 3:3450
- Heukamp LC, Thor T, Schramm A, De Preter K, Kumps C, et al. 2012. Targeted expression of mutated ALK induces neuroblastoma in transgenic mice. *Sci Transl Med* 4:141ra91
- Holliday R. 2006. Epigenetics: a historical overview. *Epigenetics* 1:76-80
- Infarinato NR, Park JH, Krytska K, Ryles HT, Sano R, et al. 2016. The ALK/ROS1 Inhibitor PF-06463922 Overcomes Primary Resistance to Crizotinib in ALK-Driven Neuroblastoma. *Cancer Discov* 6:96-107
- International HapMap Project, The. 2003 *Nature* 426:789-96
- Iwahara T, Fujimoto J, Wen D, Cupples R, Bucay N, et al. 1997. Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene* 14:439-49
- Janoueix-Lerosey I, Lequin D, Brugieres L, Ribeiro A, de Pontual L, et al. 2008. Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* 455:967-70
- Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* 293:1074-80
- Jinek M, Doudna JA. 2009. A three-dimensional view of the molecular machinery of RNA interference. *Nature* 457:405-12
- Jones PA, Laird PW. 1999. Cancer epigenetics comes of age. *Nat Genet* 21:163-7
- Kaneko Y, Kanda N, Maseki N, Sakurai M, Tsuchida Y, et al. 1987. Different karyotypic patterns in early and advanced stage neuroblastomas. *Cancer Res* 47:311-8
- Katayama R, Khan TM, Benes C, Lifshits E, Ebi H, et al. 2011. Therapeutic strategies to overcome crizotinib resistance in non-small cell lung cancers harboring the fusion oncogene EML4-ALK. *Proc Natl Acad Sci U S A* 108:7535-40
- Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, et al. 2003. Large-scale genotyping of complex DNA. *Nat Biotechnol* 21:1233-7
- Kinzler KW, Vogelstein B. 1997. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* 386:761, 3
- Knudson AG, Jr. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68:820-3
- Knudson AG, Jr., Strong LC. 1972. Mutation and cancer: neuroblastoma and pheochromocytoma. *Am J Hum Genet* 24:514-32
- Kogner P, Barbany G, Dominici C, Castello MA, Raschella G, Persson H. 1993. Coexpression of messenger RNA for TRK protooncogene and low affinity nerve



- growth factor receptor in neuroblastoma with favorable prognosis. *Cancer Res* 53:2044-50
- Krona C, Caren H, Sjöberg RM, Sandstedt B, Laureys G, et al. 2008. Analysis of neuroblastoma tumour progression; loss of PHOX2B on 4p13 and 17q gain are early events in neuroblastoma tumourigenesis. *Int J Oncol* 32:575-83
- Kushner BH, Gilbert F, Helson L. 1986. Familial neuroblastoma. Case reports, literature review, and etiologic considerations. *Cancer* 57:1887-93
- Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, et al. 2010. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 363:1693-703
- Ladenstein R, Ambros IM, Potschger U, Amann G, Urban C, et al. 2001. Prognostic significance of DNA di-tetraploidy in neuroblastoma. *Med Pediatr Oncol* 36:83-92
- Lamant L, Pulford K, Bischof D, Morris SW, Mason DY, et al. 2000. Expression of the ALK tyrosine kinase gene in neuroblastoma. *Am J Pathol* 156:1711-21
- Levan G, Mandahl N, Bregula U, Klein G, Levan A. 1976. Double minute chromosomes are not centromeric regions of the host chromosomes. *Hereditas* 83:83-90
- Li R, Morris SW. 2008. Development of anaplastic lymphoma kinase (ALK) small-molecule inhibitors for cancer therapy. *Med Res Rev* 28:372-412
- Liu Z, Thiele CJ. 2012. ALK and MYCN: when two oncogenes are better than one. *Cancer Cell* 21:325-6
- Lu J, Guan S, Zhao Y, Yu Y, Woodfield SE, et al. 2017. The second-generation ALK inhibitor alectinib effectively induces apoptosis in human neuroblastoma cells and inhibits tumor growth in a TH-MYCN transgenic neuroblastoma mouse model. *Cancer Lett* 400:61-8
- Maris JM, Guo C, Blake D, White PS, Hogarty MD, et al. 2001. Comprehensive analysis of chromosome 1p deletions in neuroblastoma. *Med Pediatr Oncol* 36:32-6
- Maris JM, Matthay KK. 1999. Molecular biology of neuroblastoma. *J Clin Oncol* 17:2264-79
- Maris JM, Weiss MJ, Guo C, Gerbing RB, Stram DC, et al. 2000. Loss of heterozygosity at 1p36 independently predicts for disease progression but not decreased overall survival probability in neuroblastoma patients: a Children's Cancer Group study. *J Clin Oncol* 18:1888-99
- Marsilje TH, Pei W, Chen B, Lu W, Uno T, et al. 2013. Synthesis, structure-activity relationships, and in vivo efficacy of the novel potent and selective anaplastic lymphoma kinase (ALK) inhibitor 5-chloro-N2-(2-isopropoxy-5-methyl-4-(piperidin-4-yl)phenyl)-N4-(2-(isopropylsulfonyl)phenyl)pyrimidine-2,4-diamine (LDK378) currently in phase 1 and phase 2 clinical trials. *J Med Chem* 56:5675-90
- Martinsson T, Sjöberg RM, Hedborg F, Kogner P. 1995. Deletion of chromosome 1p loci and microsatellite instability in neuroblastomas analyzed with short-tandem repeat polymorphisms. *Cancer Res* 55:5681-6
- Metzker ML. 2010. Sequencing technologies - the next generation. *Nat Rev Genet* 11:31-46
- Miyake I, Hakomori Y, Shinohara A, Gamou T, Saito M, et al. 2002. Activation of anaplastic lymphoma kinase is responsible for hyperphosphorylation of ShcC in neuroblastoma cell lines. *Oncogene* 21:5823-34
- Monclair T, Brodeur GM, Ambros PF, Brisse HJ, Cecchetto G, et al. 2009. The International Neuroblastoma Risk Group (INRG) staging system: an INRG Task Force report. *J Clin Oncol* 27:298-303
- Moorthie S, Mattocks CJ, Wright CF. 2011. Review of massively parallel DNA sequencing technologies. *Hugo J* 5:1-12
- Morris SW, Kirstein MN, Valentine MB, Dittmer KG, Shapiro DN, et al. 1994. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 263:1281-4
- Mosse YP, Laudenslager M, Khazi D, Carlisle AJ, Winter CL, et al. 2004. Germline PHOX2B mutation in hereditary neuroblastoma. *Am J Hum Genet* 75:727-30

- Mosse YP, Laudenslager M, Longo L, Cole KA, Wood A, et al. 2008. Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* 455:930-5
- Mosse YP, Lim MS, Voss SD, Wilner K, Ruffner K, et al. 2013. Safety and activity of crizotinib for paediatric patients with refractory solid tumours or anaplastic large-cell lymphoma: a Children's Oncology Group phase 1 consortium study. *Lancet Oncol* 14:472-80
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51 Pt 1:263-73
- Nakagawara A, Arima M, Azar CG, Scavarda NJ, Brodeur GM. 1992. Inverse relationship between trk expression and N-myc amplification in human neuroblastomas. *Cancer Res* 52:1364-8
- Nakagawara A, Azar CG, Scavarda NJ, Brodeur GM. 1994. Expression and function of TRK-B and BDNF in human neuroblastomas. *Mol Cell Biol* 14:759-67
- Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, et al. 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res* 39:e90
- Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, et al. 2005. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 65:6071-9
- Osajima-Hakomori Y, Miyake I, Ohira M, Nakagawara A, Nakagawa A, Sakai R. 2005. Biological role of anaplastic lymphoma kinase in neuroblastoma. *Am J Pathol* 167:213-22
- Passariello A, De Brasi D, Defferrari R, Genesio R, Tufano M, et al. 2013. Constitutional 11q14-q22 chromosome deletion syndrome in a child with neuroblastoma MYCN single copy. *Eur J Med Genet* 56:626-34
- Pulford K, Lamant L, Morris SW, Butler LH, Wood KM, et al. 1997. Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)-ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. *Blood* 89:1394-404
- Pulford K, Morris SW, Turturro F. 2004. Anaplastic lymphoma kinase proteins in growth control and cancer. *J Cell Physiol* 199:330-58
- Reshetnyak AV, Murray PB, Shi X, Mo ES, Mohanty J, et al. 2015. Augmentor alpha and beta (FAM150) are ligands of the receptor tyrosine kinases ALK and LTK: Hierarchy and specificity of ligand-receptor interactions. *Proc Natl Acad Sci U S A* 112:15862-7
- Ryden M, Sehgal R, Dominici C, Schilling FH, Ibanez CF, Kogner P. 1996. Expression of mRNA for the neurotrophin receptor trkC in neuroblastomas with favourable tumour stage and good prognosis. *Br J Cancer* 74:773-9
- Sakai T, Toguchida J, Ohtani N, Yandell DW, Rapaport JM, Dryja TP. 1991. Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am J Hum Genet* 48:880-8
- Sato-Otsubo A, Sanada M, Ogawa S. 2012. Single-nucleotide polymorphism array karyotyping in clinical practice: where, when, and how? *Semin Oncol* 39:13-25
- Savelyeva L, Corvi R, Schwab M. 1994. Translocation involving 1p and 17q is a recurrent genetic alteration of human neuroblastoma cells. *Am J Hum Genet* 55:334-40
- Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, et al. 2007. Challenges and standards in integrating surveys of structural variation. *Nat Genet* 39:S7-15
- Schleiermacher G, Janoueix-Lerosey I, Ribeiro A, Klijanienko J, Couturier J, et al. 2010. Accumulation of segmental alterations determines progression in neuroblastoma. *J Clin Oncol* 28:3122-30
- Schleiermacher G, Javanmardi N, Bernard V, Leroy Q, Cappo J, et al. 2014. Emergence of new ALK mutations at relapse of neuroblastoma. *J Clin Oncol* 32:2727-34

- Schleiermacher G, Raynal V, Janoueix-Lerosey I, Combaret V, Aurias A, Delattre O. 2004. Variety and complexity of chromosome 17 translocations in neuroblastoma. *Genes Chromosomes Cancer* 39:143-50
- Schonherr C, Ruuth K, Kamaraj S, Wang CL, Yang HL, et al. 2012. Anaplastic Lymphoma Kinase (ALK) regulates initiation of transcription of MYCN in neuroblastoma cells. *Oncogene* 31:5193-200
- Schroder J, Bailey J, Conway T, Zobel J. 2010. Reference-free validation of short read data. *PLoS One* 5:e12681
- Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, et al. 1983. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* 305:245-8
- Shaw AT, Kim DW, Mehra R, Tan DS, Felip E, et al. 2014. Ceritinib in ALK-rearranged non-small-cell lung cancer. *N Engl J Med* 370:1189-97
- Shiota M, Fujimoto J, Semba T, Satoh H, Yamamoto T, Mori S. 1994. Hyperphosphorylation of a novel 80 kDa protein-tyrosine kinase similar to Ltk in a human Ki-1 lymphoma cell line, AMS3. *Oncogene* 9:1567-74
- Siaw JT, Wan H, Pfeifer K, Rivera VM, Guan J, et al. 2016. Brigatinib, an anaplastic lymphoma kinase inhibitor, abrogates activity and growth in ALK-positive neuroblastoma cells, Drosophila and mice. *Oncotarget* 7:29011-22
- Speicher MR, Carter NP. 2005. The new cytogenetics: blurring the boundaries with molecular biology. *Nat Rev Genet* 6:782-92
- Srivatsan ES, Ying KL, Seeger RC. 1993. Deletion of chromosome 11 and of 14q sequences in neuroblastoma. *Genes Chromosomes Cancer* 7:32-7
- Sung KW. 2012. Treatment of high-risk neuroblastoma. *Korean J Pediatr* 55:115-20
- Suzuki T, Bogenmann E, Shimada H, Stram D, Seeger RC. 1993. Lack of high-affinity nerve growth factor receptors in aggressive neuroblastomas. *J Natl Cancer Inst* 85:377-84
- Suzuki T, Yokota J, Mugishima H, Okabe I, Ookuni M, et al. 1989. Frequent loss of heterozygosity on chromosome 14q in neuroblastoma. *Cancer Res* 49:1095-8
- Thompson PM, Seifried BA, Kyemba SK, Jensen SJ, Guo C, et al. 2001. Loss of heterozygosity for chromosome 14q in neuroblastoma. *Med Pediatr Oncol* 36:28-31
- Trochet D, Bourdeaut F, Janoueix-Lerosey I, Deville A, de Pontual L, et al. 2004. Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. *Am J Hum Genet* 74:761-4
- Valton AL, Dekker J. 2016. TAD disruption as oncogenic driver. *Curr Opin Genet Dev* 36:34-40
- Watson JD, Crick FH. 1953. The structure of DNA. *Cold Spring Harb Symp Quant Biol* 18:123-31
- Yamamoto G, Nannya Y, Kato M, Sanada M, Levine RL, et al. 2007. Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am J Hum Genet* 81:114-26
- Yamashiro DJ, Nakagawara A, Ikegaki N, Liu XG, Brodeur GM. 1996. Expression of TrkC in favorable human neuroblastomas. *Oncogene* 12:37-41
- Zhao X, Li C, Paez JG, Chin K, Janne PA, et al. 2004. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 64:3060-71
- Zhu S, Lee JS, Guo F, Shin J, Perez-Atayde AR, et al. 2012. Activated ALK collaborates with MYCN in neuroblastoma pathogenesis. *Cancer Cell* 21:362-73