INSIGHTS INTO NEUROBLASTOMA INITIATION AND DISEASE PROGRESSION THROUGH INTEGRATIVE GENOMICS AND EPIGENOMICS

Derek Alan Oldridge

A DISSERTATION

in

Genomics and Computational Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

_____

John M. Maris

Professor of Pediatrics

Co-Supervisor of Dissertation

_____

Sharon J. Diskin

Assistant Professor of Pediatrics

Graduate Group Chairperson

_____

Li-San Wang, Associate Professor of Pathology and Laboratory Medicine

Dissertation Committee

Chair: Lewis A. Chodosh, Professor of Cancer Biology

Christopher D. Brown, Assistant Professor of Genetics

Andrea Califano, Professor of Chemical Systems Biology

John B. Hogenesch, Professor of Pharmacology

Shane T. Jensen, Associate Professor of Statistics

INSIGHTS INTO NEUROBLASTOMA INITIATION AND DISEASE PROGRESSION THROUGH INTEGRATIVE GENOMICS AND EPIGENOMICS

Acknowledgements

Everything in this dissertation I owe to the phenomenal mentorship of my advisor and co-advisor, John Maris and Sharon Diskin. As an exemplary physician-scientist who has dedicated his life to improving outcomes for children with cancer, I can only hope to emulate John's depth of clinical insight, intelligence, and intuition for choosing the most impactful scientific questions to pursue and answering them with rigor and tenacity. And as a aerospace-engineer turned pediatric cancer biologist, Sharon's career exemplifies how quantitative and computational approaches are transforming biology, providing me an invaluable blueprint for how to have success in my own career in the years ahead. Thank you both for your guidance and support these past few years and in anticipation of the years to come, not to mention the countless answered emails, unscheduled office chats, recommendations, and providing me many opportunities to network with the broader cancer research community and establish myself as a young scientist.

Thank you to Lewis Chodosh, for serving as my thesis committee chair, and to the rest of my committee, Casey Brown, Andrea Califano, John Hogenesch, and Shane Jensen, for taking the time out of your busy schedules, for many thought-provoking conversations, and for pushing me to be the best researcher I can be. In particular, thank you John Hogenesch for our early conversations in GCB531: Introduction to Genome Sciences, which provided the initial spark to utilize neuroblastoma ENCODE data to map causal variants and formed the basis of the LMO1 causal SNP project.

Thank you to the all the members of the Maris and Diskin laboratories for providing a fun and stimulating environment in which to train and for your many intellectual and experimental contributions to the work presented here. In particular, thank you to Andy Wood for your collaboration on our shared LMO1 project and our for your mentorship

as a physician-scientist. Thank you Ian Crimmins, Cindy Winter, Robyn Sussman, Lori Soffer, Julie-Ann Rader, Edward Attiyeh, Lee McDaniel, Karina Conkrite, Perry Evans, and Tom Watkins for all of your scientific contributions to this work, for all you have taught me, and for always keeping lab work entertaining!

The projects described here benefitted greatly from the input of many external collaborators. Thank you Nina Weichert-Leahey, Jane Zhu, Adam Durbin, Brian Abraham, Takaomi Sanda, Mario Capasso, Nazneen Rahman, Rick Young, Tom Look, and Gerd Blobel and for your invaluable discussions and experimental contributions to our LMO1 causal SNP project. Thank you Thomas Eleveld, Virginie Bernard, Jan Koster, Olivier Delattre, Rogier Versteeg, Gudrun Schleiermacher, and Jan Molenaar: it was a pleasure meeting many of you at the Advances in Neuroblastoma Research 2014 Congress and merging our collaborative efforts to understand the molecular basis of relapsed neuroblastoma. Thank you also to all the members of the TARGET Project, who provided a large portion of the next-generation sequencing data used here.

Thank you to the Genomics and Computational Biology graduate group and broader Penn community for providing a fantastic training environment in which to pursue my PhD. Thank you Maja Bucan and Lisan Wang for your stewardship of GCB and Skip Brass and Maggie Krall for running the best MSTP in the Galaxy. Thank you for all of your advice and support throughout my training. Thank you Hannah Chervitz and Maureen Kirsch for your tireless work to keep our program running smoothly.

Thank you to the Wharton Statistics Department for taking me in as one of your own during my time as a graduate student. Thank you in particular to Andreas Buja for sparking my interest in statistics, and to Shane Jensen, Dylan Small, and Larry Brown for sustaining it over three years of coursework. Thank you Nancy Zhang,

Yuchao Jiang, Sharon Diskin, Andreas Buja and Dylan Small for your support when I decided to add a statistics master's degree to my graduate training.

Thank you also to countless teachers and mentors who nurtured my academic and scientific interests from a young age and to my Reed College professors who set me on my current research path. Thank you to Jay Mellies, for taking me on even as a freshman and getting me hooked on research; to David Griffiths and Johnny Powell, for convincing me that the best way to prepare myself for a career in biology was to major in physics; to Alan Shusterman, for getting me hooked on computational research; and to Arthur Glasfeld, who taught me everything I know about structural biochemistry and who would be pleased to see I haven't veered too far from transcription factor biology. Thank you Massimo Loda, Francesca Demichelis, Mark Rubin, and Vivian Cheung for your mentorship in oncology and in genomics research, setting the stage for my PhD work here. Thank you Debra Leonard, Robert Wilson, Steve Master, and Roger Greenberg for your continuing career mentorship as physician-scientists.

Thank you to GCB and the National Cancer Institute for your financial support of this work through a T32 training grant and F30 individual fellowship. Thank you to the Schweisguth Prize selection committee for your recognition of my LMO1 causal SNP work and for the invitation to speak at the International Society of Pediatric Oncology 2015 Annual Conference. Thank you to the Department of Genetics and the 2015 Kadesch Prize selection committee for your recognition of my dissertation work, it has been a great honor.

Finally, thank you to my friends and family for your continuing love and support, and especially to my Mom for her incredible strength and patience, for raising me so well, and for introducing me to computers at a young age. Last but not least, thank you Olivia Padovan-Merhar for being my best friend and companion for nine years and counting. I look forward to many more.

ABSTRACT

INSIGHTS INTO NEUROBLASTOMA INITIATION AND DISEASE PROGRESSION THROUGH INTEGRATIVE GENOMICS AND EPIGENOMICS

Derek A. Oldridge

John M. Maris

Sharon J. Diskin

In this dissertation, we use integrative genomics to shed new insights into the molecular lesions and mechanisms that drive neuroblastoma. In Part 1, we use imputation and epigenetic profiling in order to identify the causal germline SNP that drives differential susceptibility to neuroblastoma at the *LMO1* oncogene locus. In Part 2, we use whole genome sequencing and Bayesian statistical modeling to understand the clonal evolution that occurs between diagnosis and relapse.

Part 1: Neuroblastoma is a pediatric malignancy that typically arises in early childhood, and is derived from the developing sympathetic nervous system. A previous genome-wide association study identified common polymorphisms at the *LMO1* gene locus that are highly associated with neuroblastoma susceptibility and oncogenic addiction to LMO1 in the tumor cells. Here we investigate the causal DNA variant at this locus. We show that SNP rs2168101 G→T is the most highly associated variant and resides in a super-enhancer defined by extensive acetylation of histone H3 lysine 27 within the first intron of *LMO1*. The ancestral G allele that is associated with tumor formation resides in a conserved GATA transcription factor binding motif. We show that the newly evolved protective TATA allele ablates GATA3 binding and enhancer activity, and is associated with decreased total and allele-specific *LMO1* expression in neuroblastoma primary tumors. These findings indicate that a recently evolved polymorphism within a super-enhancer element in the first intron of *LMO1* influences

neuroblastoma susceptibility through differential GATA transcription factor binding and direct modulation of *LMO1* expression *in cis*.

Part 2: The majority of high-risk neuroblastomas initially respond to chemotherapy, but over half of patients will experience therapy-resistant relapses which are nearly always fatal. The molecular defects driving relapse and drug resistance are unknown. We performed whole genome sequencing of 23 paired diagnostic and relapsed neuroblastomas, and corresponding normal lymphocyte DNAs, to define genetic alterations associated with relapse. Unbiased pathway analysis of the somatic mutations detected in the relapse tissues identified a strong enrichment in genes associated with RAS-MAPK signaling (18 of 23 patients). These RAS-MAPK mutations were clonally enriched at relapse and exist within clonal or major subclonal tumor populations. Similar MAPK pathway mutations were detected in 11 of 18 human neuroblastoma-derived cell lines, and these lesions are predicted to be sensitive to small molecule inhibition of MEK *in vitro* and *in vivo*. In this study of 23 neuroblastoma cases, MAPK pathway mutations were highly enriched in the relapsed genomes, providing a potential biomarker for new therapeutic approaches to chemotherapy refractory disease.

Collectively, these studies provide important insights into the genetic and epigenetic factors driving neuroblastoma, and suggest new opportunities for pathway-targeted therapies.

Contents

List of Tables

List of Figures

Chapter 1

An introduction to neuroblastoma genomics.

1.1 Neuroblastoma at a glance and the promise of targeted therapies

Neuroblastoma is a neural-crest-derived cancer of the developing sympathetic nervous system that most commonly arises from the adrenal gland of children under 5 years of age, but can present anywhere along the sympathetic chain[1, 2]. It accounts for 10% of all pediatric cancer and is responsible for close to 15% of all pediatric cancer deaths[1, 2]. Although familial neuroblastoma with an autosomal dominant inheritance pattern occurs in approximately 1-2% of all cases, the majority of neuroblastomas are sporadic[1, 2]. Moreover, neuroblastoma is recognized as a highly heterogeneous malignancy that includes multiple clinically and molecularly distinct subclasses[3–5]. Representing one extreme is stage 4S neuroblastoma, which presents as widely disseminated disease, but is characterized by the possibility of spontaneous regression and survival probability of 92%[2, 6]. On another extreme is high-risk neuroblastoma, which is characterized by MYCN gene amplification, more advanced stage, and/or older age of onset[7]. The prognosis for high-risk neuroblastoma remains poor with a 5

year survival probability of approximately 40%, and relapsed high-risk neuroblastoma is nearly always fatal, with a 5 year overall survival rate of less than 5%[1, 2]. These poor outcomes persist in spite of aggressive treatment typically consisting of surgery, radiotherapy, and/or cytotoxic chemotherapy, as well as the recent addition of anti-GD2 immunotherapy[8]. Moreover, even when efficacious, such treatments can leave patients with long-term physical or cognitive impairment[9].

The past two decades have seen the emergence of a generation of cancer therapeutics that directly target the underlying molecular drivers of specific malignancies, holding great promise to increase treatment efficacy while simultaneously decreasing treatment-related toxicity[10]. The development of trastuzumab for treatment of HER2-amplified breast cancer[11] and imatinib for BCR-ABL-translocated chronic myelogenous leukemia[12] are among the earliest, prototypical examples of how identification of molecular lesions has been exploited for therapeutic benefit in specific cancers. This paradigm, successfully applied in other cancers, provides the rationale and impetus for characterizing the genome of neuroblastoma: first, to understand the fundamental molecular basis of this disease, and ultimately to translate this understanding into better treatments. As immune checkpoint blockade and other immune-based therapies have recently emerged as new and powerful treatment modalities[13, 14], genomic profiling may take on additional importance in neo-antigen prediction and in identifying patients who may benefit from the combination of targeted small molecule therapies and new immunotherapies[15].

1.2 Neuroblastoma genetics in the pre-genomics era

The first insights into the somatic genome of neuroblastoma came from karyotyping studies, which revealed recurrent broad segmental chromosomal rearrangements—including 1p deletion, 11q deletion, and 17q gain—that have long been appreciated for their prognostic relevance[16–20]. The identification of the first *bone fide* neuroblastoma oncogene came in the discovery a c-MYC homologue (named N-Myc or MYCN, for 'N'euroblastoma) that manifested as double minute chromosomes or homogeneously staining regions of chromosomes—reflecting $>$ 100-fold amplications of the *MYCN* oncogene—in a subset of neuroblastoma tumors[21]. It is now appreciated that high copy MYCN-amplifications occur in approximately 20% of all neuroblastomas and are associated with significantly worse prognosis[22]. Indeed, *MYCN*-amplificiation is one of the primary criteria for determining what constitutes "high-risk" neuroblastoma clinically[7].

The earliest efforts to characterize somatic point mutations in neuroblastoma, informed as they were by common mutation types that had been observed in other cancers, were met with relatively little success. For example, while p53 tumor suppressor mutations are estimated to occur in approximately half of all cancers[23], it was discovered early on that p53 mutations are very rare in most neuroblastomas[24, 25]. Similar targeted sequencing studies revealed that RAS oncogene mutations, common in several other cancer types[26, 27], were also quite rare in neuroblastoma[28, 29]. However, such candidate gene sequencing approaches did meet early luck in the discovery of *PHOX2B* as a familial neuroblastoma gene[30, 31], which was motivated by earlier discovery of *PHOX2B* mutation as a major cause of another disease of the neural crest: Congenital Central Hypoventilation Syndrome, also known as "Ondine's

Curse"[32]. In aggregate, the genes identified by these early studies had identified key oncogenic driver mutations for only a small fraction of neuroblastomas, leaving open the possibility that neuroblastoma drivers might be hidden in some as yet uncharted region of the genome and that their discovery would need to await the arrival of technologies that enabled just such an unbiased, genome-wide search.

1.3 Neuroblastoma genetics in the genomics era

The completion of the Human Genome Project[33, 34], and with it the emergence of microarray and next generation sequencing (NGS) technologies[35] opened new avenues for understanding neuroblastoma genetics at a genomic scale. An important breakthrough came in 2008, when an unbiased linkage analysis uncovered germline mutations in the kinase domain of Anaplastic Lymphoma Kinase (*ALK*) as the major cause of familial neuroblastoma[36]. Concurrently, genome-wide copy number profiling had identified high copy somatic *ALK* amplifications in a subset of neuroblastoma tumors, and subsequent resequencing demonstrated that point mutations in *ALK* occur in approximately 10% of all neuroblastomas[36–39]. In a stroke of serendipity, hyper-active ALK-signaling mediated by *EML4-ALK* translocations had already been identified as a common and druggable lesion in Non-Small Cell Lung Cancer, with an ALK-inhibitor already in active development[40, 41]. Ultimately, these converging lines of evidence uncovered ALK as an important driver in a substantial fraction of neuroblastomas, and the translation of this discovery into targeted therapies for ALK-mutant neuroblastoma remains an exciting and active area of study[42, 43].

As the cost of whole exome sequencing (WES) and whole genome sequencing (WGS) have continued to decline, an unbiased picture of the landscape of the neuroblastoma protein coding genome has emerged over the past few years[44–47]. In 2012, our Dutch

colleagues published a study of 87 neuroblastoma tumors of all stages, which aside from alterations in *MYCN* and *ALK*, identified recurrent structural alterations in *ODZ3* (5.7%), *PTPRD* (5.7%) and *CSMD1* (3.4%)—postulated as regulators neuronal growth cone stabilization—as well as recurrent structural alterations or point mutations in *ATRX* (5.7%), *TIAM1* (3.4%), and other regulators of the Rac/Rho pathway[44]. Our laboratory leads the neuroblastoma component of the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) pediatric cancer genomics project, and in 2013 published a whole exome sequencing study of 222 tumors also showing a relative dearth of somatic coding mutations, with the most frequently mutated genes in neuroblastoma being *ALK* (9%), *PTPN11* (2.9%), *ATRX* (2.5%, and an additional 7.1% had focal deletions), and *NRAS* (0.83%)[46]. Unfortunately, the relative lack of recurrent somatic mutations found in these studies raises important doubts as to whether precision-based medicine targeted to mutations detected in diagnostic tumors can be readily generalized to the majority of neuroblastomas.

While the idea that common germline variation might contribute to sporadic neuroblastoma was initially met with skepticism, an ongoing genome wide association study (GWAS) has definitively established that common variation contributes to both neuroblastoma tumor initiation and maintenance, implicating many additional genes in neuroblastoma tumoriogenesis. These include variants in or near the genes of *BARD1*[48], *CASC15*[49], *LMO1*[50], *LIN28B*[51], *HACE1*[51], *TP53*[52], associated with high-risk neuroblastoma; variants in or near *DUSP12*, *DDX4*, and *HSD17B12*[53], associated with low-risk neuroblastoma; and copy number variation resulting in deletion of the *NBPF23* gene[54]. Although many of these genes have been previously implicated in other cancers and quite plausibly function as neuroblastoma oncogenes or tumor suppressors, much work remains to be done to better understand the mechanism for how genetic variation affects neuroblastoma and to understand how these genes

function in neuroblastoma pathogenesis more broadly.

1.4 Summary of challenges in the field of neuroblastoma genomics, circa 2012

The previous sections summarize what was understood in the field of neuroblastoma genomics prior to the start of my Ph.D. training. On the one hand, an exciting period of discovery had just implicated neuroblastoma-associated common germline variation near genes previously unknown to affect disease-initation, but the causal mechanisms remained to be elucidated. On the other hand, the paucity of recurrent somatic coding mutations in neuroblastoma provided no clear strategy for the development of rational therapeutics, with the notable exception of *ALK*-mutated cases. Moreover, essentially nothing was known about how neuroblastomas tumors evolve over the course of therapy, nor what molecular lesions underlie the tumors of patients who ultimately relapse and who remain in dire need of effective treatment strategies. In the following chapters, I will summarize how my dissertation work has begun to shed light on exactly these issues.

Chapter 2

Non-coding mechanisms of functional dysregulation in neuroblastoma

2.1 Introduction to regulatory genetics and epigenetics

Gene expression programs responsible for cellular differentiation and function are highly dependent on the packaging of chromatin, a macromolecular complex of DNA and DNA-binding histone proteins[55]. The partitioning of chromatin into functional domains is largely determined by chemical, "epigenetic" alterations to DNA and histones that do not affect the underlying nucleic acid sequence, and by DNA-binding protein complexes that mediate folding of the genome by pulling together distal regions of DNA polymers into close proximity in three-dimensional space[56]. This folding, in turn, establishes which genomic regions will be tightly packaged—transcriptionally inert "heterochromatin"—versus those regions which will be open and accessible to the transcriptional machinery—transcriptionally active "euchromatin". This essential partitioning of a genome into transcriptionally active and inactive domains is thought

to play fundamental roles in normal cellular development, differentiation, as well as disease[57], and provides an explanation for how cells with largely the same genetic code can be molecularly, morphologically, and functionally distinct[56]. Whereas the Human Genome Project can be understood as providing the first one-dimensional structure of the human genome, genome science is increasingly focused on solving the three-dimensional structure of nuclear DNA and finding its functional and pathological correlates[57–59].

The paucity of recurrent coding mutations in neuroblastoma suggests that non-coding lesions, including DNA mutations and epigenetic changes, may act as drivers in a substantial subset of neuroblastoma tumors. Indeed, reports of highly recurrent *TERT* promoter mutations driving oncogenic telomerase overexpression in melanoma through the de novo generation of an ETS transcription factor binding site illustrate the importance of non-coding mechanisms of cellular dysregulation in cancer[60, 61]. Additionally, the release of Encyclopedia Of DNA Elements (ENCODE)[62] and Roadmap Epigenomics[63] data—encompassing coding and non-coding RNA, transcription factor binding, and chromatin profiling in a wide variety of normal and diseased tissues—provides an unprecedented opportunity to investigate how germline variation and somatic mutations within newly identified non-coding regulatory regions may impact cancer.

In meta-analyses of all genotype-phenotype associations discovered to date through GWAS, is has been estimated that almost 90% of all disease- or phenotype-associated variants may fall within non-coding regions of the genome[64, 65], raising the important question of how genetic variation is able to affect phenotypes without changing protein coding sequences. It is now appreciated that many of these non-coding variants fall within regulatory regions known as enhancers—regulatory regions that bind to transcription factors and help promote transcription by looping to distal promoters—

and that these regulatory variants modulate enhancer activity (and hence target-gene expression) by disrupting or generating *de novo* transcription factor binding sites[65]. Although initially identified in other diseases, this paradigm is emerging as an important determinant of gene pathway dysregulation in cancer[66]. For example, a causal germline polymorphism impacting predisposition to breast, prostate, and colon cancer was discovered to affect MYC expression through modulation of TCF transcription factor binding within an enhancer region upstream of the *MYC* proto-oncogene[67, 68]. In some adult cancers where mature GWAS efforts have identified many cancer-association loci, the focus is now turning to functionally validating many regions in parallel[69, 70].

Uncovering causal variants from GWAS is difficult for many reasons. One challenge is posed by the high degree of linkage disequilibrium in the human genome, resulting in strong correlation and statistical indistinguishability between neighboring variants[71]. Additionally, high-resolution chromatin maps are largely incomplete for most tissues, whereas mapping needs to be performed in the disease-relevant tissue or cell line due to the highly tissue-specific nature of chromatin structure. However, of all the neuroblastoma susceptibility loci that were been identified to as of 2012, the *LMO1* oncogene was an especially attractive candidate for searching for regulatory variant candidates. First, in the initial 2011 study implicating the *LMO1* locus in neuroblastoma, it was observed that risk alleles were associated with higher gene expression[50], consistent with a regulatory phenotype. Second, only neuroblastoma cell lines possessing the risk allele showed sensitivity to *LMO1* knockdown[50], indicating that *LMO1* expression played an allele-dependent role in tumor maintenance in addition to initiation, thereby suggesting that neuroblastoma cell lines could be useful models in uncovering causal variants. Finally, the publication of ENCODE data at the end of 2012[62], which included DNase I hypersensitivity sequencing marking active enhancer regions along

with binding profiles of an array of other transcription factors in neuroblastoma cell lines, provided the perfect opportunity to see if the regulatory architecture of the *LMO1* locus could help identify candidate functional variants. The following section describes what we found, and is largely copied with minor edits from our recent publication in *Nature*, whose full citation is provided here: [72]. Although this study represents the collaborative efforts of many people, my principal contributions included the conceptualization and execution of the integration of ENCODE-related data that identified rs2168101 as a causal variant candidate, performing nearly all bioinformatic analyses excluding genome-wide imputation and super-enhancer analysis, helping guide the experimental validation efforts which included performing reporter assays and ChIP-PCR experiments myself, as well as writing the majority of the paper and making nearly all of the figures.

2.2 Oldridge, Wood, *et al.* Genetic predisposition to neuroblastoma mediated by a *LMO1* super-enhancer polymorphism.

Genome-wide association study (GWAS) efforts frequently identify highly statistically significant genetic associations within non-coding regulatory regions of the genome, but the underlying causal DNA sequence variations have only been identified in a few instances. A neuroblastoma GWAS has identified several disease susceptibility loci[48–54], with the signal within the LIM domain only 1 (*LMO1*) locus at 11p15[50], a transcriptional co-regulator containing two zinc finger LIM domains that nucleate and regulate transcription factor complexes, being most robust. The main members of the LMO gene family, LMO1-4, are all implicated in cancer including *LMO1* and

LMO2 translocations in T-cell leukemia[73], and we previously provided the first evidence that *LMO1* was a *bona fide* neuroblastoma oncogene[50]. Here, we sought to identify the causal polymorphism(s) driving the *LMO1* genetic association with neuroblastoma susceptibility as a basis for understanding neuroblastoma initiation and addiction mechanisms.

We first performed fine mapping of associated germline SNPs and indels at the *LMO1* gene locus by imputation to the 1000 Genomes Project for our European-American neuroblastoma GWAS[51]. This identified 27 SNPs with minor allele frequency (MAF) > 0.01 and an association $P < 1 \times 10^{-5}$ (Figure 2.1 and Table 2.1). We further prioritized associated variants by evolutionary conservation, and by their regulatory potential inferred through neuroblastoma-specific DNase I hypersensitivity mapping and chromatin immunoprecipitation sequencing (ChIP-seq) from the ENCODE Consortium (Figure 2.2). These data showed that the most significantly associated SNP at the *LMO1* locus (rs2168101, odds ratio = 0.67, $P = 4.14 \times 10^{-16}$) resides within a highly conserved and active enhancer region inferred by DNase I sensitivity and p300 binding in the SKNSH neuroblastoma cell line (Figure 2.2). Notably, we found no rare or common non-synonymous coding variants in *LMO1* in a combined cohort of 482 unique neuroblastoma cases with germline whole-genome ($n = 136$), whole-exome ($n = 222$) and/or targeted DNA sequencing ($n = 183$) (see Table 2.2).

Because rs2168101 genotypes were imputed in our analyses (Figure 2.3), we next directly genotyped this SNP in 146 out of 2,101 European-American cases, and measured an 86% imputation accuracy. We additionally directly genotyped rs2168101 in two independent cohorts from the UK and Italy, with both showing robust replication (Table 2.3). We did not observe replication in an independent African-American cohort. Notably, the protective T allele is common in Europeans (CEU HapMap: 28%) and

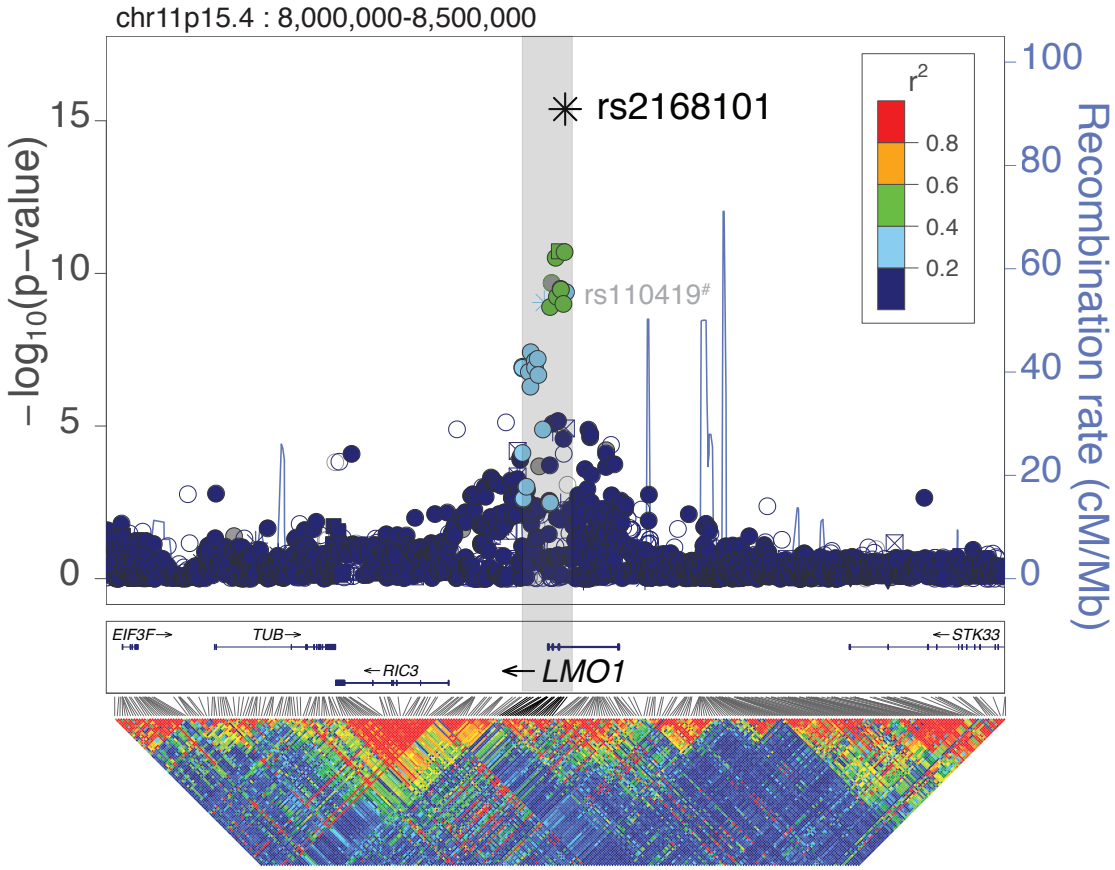

Figure 2.1: Imputation GWAS identifies additional neuroblastoma-associatied variants at the *LMO1* locus. Manhattan plot for neuroblastoma GWAS (cases = 2,101; controls = 4,202). The neuroblastoma-associated region falls within a 40-kilobase (kb) haplotype block (grey box) in Europeans, encompassing the *LMO1* 3'-terminus. rs2168101 is the most associated variant and is moderately correlated (maximum $r^2 = 0.52$) with other variants. The sentinel SNP reported previously, rs110419, is also highlighted (#).

East Asians (CHB+JPT HapMap: 32%), but is rare or absent in Africans, indicating recent expansion of the rs2168101 protective allele in non-African human populations (Figure 2.4). Meta-analysis demonstrated a combined association $P = 7.47 \times 10^{-29}$ across 8,553 controls and 3,254 cases (Table 2.3).

As causal SNPs driving GWAS associations may disrupt transcription factor binding

Table 2.1: Germline variants from 1000 Genomes Project associated with neuroblastoma susceptibility ($P < 10^{-5}$) in European-American cohort from imputation-based analysis.

Variant ID (rsID)	Chrom (hg19)	Position (hg19)	Alleles (Ref/Alt)†	Alt Allele Freq (Cases)†	Alt Allele Freq (Controls)†	*P*-Value (Additive)‡	Odds Ratio (Additive)‡
rs191871553	11	8222464	C/T	0.035 (n=2101)	0.054 (n=4202)	7.49×10^{-06}	0.64 (0.53-0.78)
rs11041809	11	8231605	A/G	0.498 (n=2101)	0.440 (n=4202)	1.13×10^{-07}	0.80 (0.74-0.87)
rs11041811	11	8231665	C/T	0.492 (n=2101)	0.434 (n=4202)	1.28×10^{-07}	0.80 (0.74-0.87)
rs11041812	11	8231684	C/T	0.492 (n=2101)	0.433 (n=4202)	1.22×10^{-07}	0.80 (0.74-0.87)
rs11041813	11	8235207	T/C	0.478 (n=2101)	0.420 (n=4202)	1.67×10^{-07}	0.81 (0.75-0.87)
rs10839999	11	8236083	G/A	0.480 (n=2101)	0.423 (n=4202)	5.06×10^{-07}	0.81 (0.75-0.88)
rs10769885	11	8236262	C/A	0.513 (n=2101)	0.453 (n=4202)	3.77×10^{-08}	0.80 (0.74-0.87)
rs4758049	11	8238428	A/C	0.511 (n=2101)	0.452 (n=4202)	7.48×10^{-08}	0.81 (0.74-0.87)
rs4758050	11	8238545	G/C	0.511 (n=2101)	0.452 (n=4202)	7.34×10^{-08}	0.81 (0.74-0.87)
rs4758051	11	8238639	G/A	0.510 (n=2101)	0.452 (n=4202)	1.22×10^{-07}	0.81 (0.75-0.87)
rs10840000	11	8240113	G/C	0.509 (n=2101)	0.450 (n=4202)	6.22×10^{-08}	0.80 (0.74-0.87)
rs7933766	11	8240464	G/A	0.511 (n=2101)	0.453 (n=4202)	2.09×10^{-07}	0.81 (0.75-0.88)
rs11041816	11	8243798	A/G	0.397 (n=2101)	0.456 (n=4202)	8.99×10^{-10}	0.77 (0.71-0.84)
rs4315061	11	8247020	T/C	0.425 (n=2101)	0.490 (n=4202)	1.25×10^{-09}	0.78 (0.72-0.84)
rs72474792	11	8247885	TATAAAA/T	0.524 (n=2101)	0.456 (n=4202)	2.04×10^{-10}	0.77 (0.71-0.84)
rs12797723	11	8247984	C/T	0.443 (n=2101)	0.514 (n=4202)	2.05×10^{-10}	0.77 (0.71-0.84)
rs2290451	11	8248440	C/G	0.295 (n=2101)	0.255 (n=4202)	8.20×10^{-06}	1.23 (1.12-1.34)
rs7952320	11	8250143	G/C	0.408 (n=2101)	0.480 (n=4202)	3.03×10^{-11}	1.31 (1.21-1.42)
rs4758317	11	8250811	C/A	0.514 (n=2101)	0.447 (n=4202)	5.76×10^{-10}	0.78 (0.72-0.84)
rs11041820	11	8251438	G/A	0.294 (n=2101)	0.253 (n=4202)	6.77×10^{-06}	1.23 (1.12-1.34)
rs3750952	11	8251921	G/C	0.408 (n=2101)	0.481 (n=4202)	1.89×10^{-11}	0.76 (0.70-0.83)
rs110419	11	8252853	A/G	0.441 (n=2101)	0.511 (n=4202)	3.16×10^{-10}	0.78 (0.72-0.84)
rs110420	11	8253049	T/C	0.441 (n=2101)	0.511 (n=4202)	3.36×10^{-10}	0.78 (0.72-0.84)
rs204928	11	8254433	A/G	0.444 (n=2101)	0.512 (n=4202)	9.85×10^{-10}	0.78 (0.72-0.85)
rs204926	11	8255106	G/A	0.440 (n=2101)	0.510 (n=4202)	1.97×10^{-11}	0.76 (0.70-0.82)
rs2168101	11	8255408	C/A	0.242 (n=2101)	0.313 (n=4202)	4.14×10^{-16}	0.67 (0.61-0.74)
rs7948497	11	8255855	C/G	0.479 (n=2101)	0.419 (n=4202)	4.05×10^{-10}	1.30 (1.20-1.41)

†Forward strand hg19, imputed genotypes from IMPUTE2, frequencies as reported by SNPTEST.
‡SNPTEST, frequentist score test with additive model, adjusted for gender and top 20 MDS components.

at distal enhancers, we sought to identify candidate SNPs disrupting known JASPAR motifs[74], which revealed that lead candidate SNP rs2168101 resides in a highly conserved GATA-binding motif (5'-A[G/T]ATAA-3', mammalian phastCons score = 100%) (Figure 2.2). Examination of a co-crystallographic structure of GATA3 bound to its cognate binding motif[75] revealed that that arginine 276 (N-terminal zinc finger) and arginine 330 (C-terminal zinc finger) both make major groove hydrogen-bonding contacts with the guanines of separate GATA motifs; either is likely to be sterically hindered by the methyl group of a substituting thymine, providing structural insight into the preferential binding of GATA3 to the 5'-GATA-3' DNA sequence rather

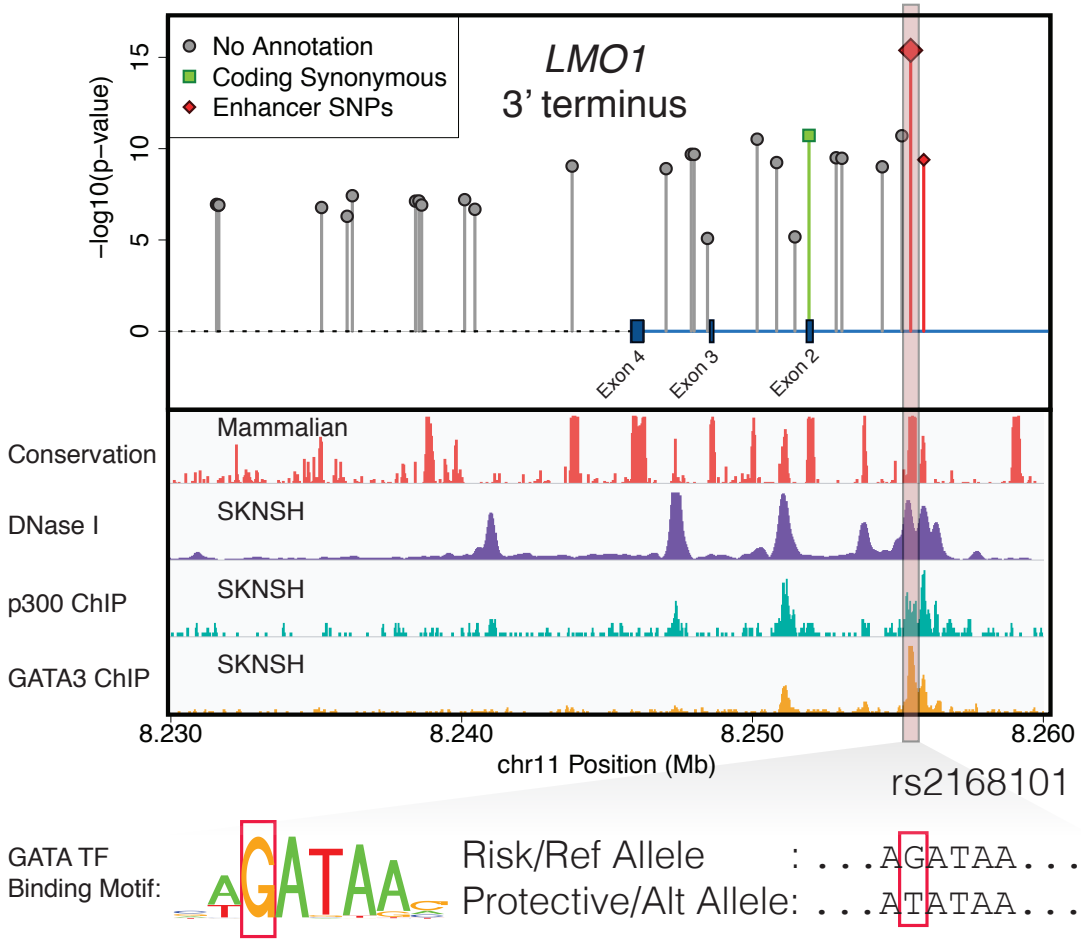

Figure 2.2: Integrative ENCODE analysis reveals that rs2168101 is an *LMO1* enhancer SNP. Neuroblastoma-associated variants ($P < 1 \times 10^{-5}$) are plotted with ENCODE tracks for neuroblastoma the cell line SKNSH. Two SNPs, rs2168101 and rs7948497, were annotated "enhancer SNPs" based on overlapping DNase peaks binding p300. The rs2168101 G→T SNP disrupts an evolutionarily conserved GATA transcription factor (TF) motif (5'-A[G/T]ATAA-3'). SKNSH has a rs2168101 = G/G genotype that preserves GATA binding, supported by ENCODE GATA3 ChIP-seq.

than 5'-TATA-3' (see Figure 2.5). ENCODE transcription factor ChIP-seq confirmed GATA2 and GATA3 binding at the rs2168101 GATA motif in the neuroblastoma cell lines SKNSH and SHSY5Y, which are G/G homozygous, thereby preserving the GATA motif (Figure 2.2). No other associated variant showed this unique combination of

Table 2.2: Patient clinical characteristics in referenced sequencing datasets.

Clinical Category	Whole Genome Seq (Blood/Tumor, $n = 136$)†	Whole Exome Seq (Blood/Tumor, $n = 222$)†	*LMO1*-Targeted Seq (Blood, $n = 183$)	Transcriptome Seq (Tumor, $n = 127$)
Age				
$<$ 18 mos	0 (0%)	32 (24%)	82 (45%)	8 (6%)
$\geq$ 18 mos	219 (100%)	103 (76%)	101 (55%)	119 (94%)
Not Available	3	1	0	0
INSS Stage‡				
Stage 1	0 (0%)	0 (0%)	39 (21%)	0 (0%)
Stage 2A	0 (0%)	0 (0%)	13 (7%)	0 (0%)
Stage 2B	0 (0%)	1 (1%)	18 (10%)	0 (0%)
Stage 3	0 (0%)	6 (4%)	27 (15%)	6 (5%)
Stage 4	219 (100%)	105 (78%)	78 (43%)	121 (95%)
Stage 4S	0 (0%)	23 (17%)	8 (4%)	0 (0%)
Not Available	3	1	0	0
MYCN				
Not Amplified	143 (67%)	102 (76%)	151 (83%)	95 (75%)
Amplified	71 (33%)	32 (24%)	30 (17%)	31 (25%)
Not Available	8	2	2	1
Histology				
Favorable	4 (2%)	29 (23%)	95 (54%)	9 (8%)
Unfavorable	187 (98%)	96 (77%)	82 (46%)	107 (92%)
Not Available	31	11	6	11
DNA Index				
Hyperdiploid	117 (54%)	81 (61%)	121 (67%)	67 (53%)
Diploid	98 (46%)	52 (39%)	59 (33%)	59 (47%)
Not Available	7	3	3	1
Risk				
Low	0 (0%)	15 (11%)	64 (35%)	0 (0%)
Intermediate	0 (0%)	14 (10%)	49 (27%)	6 (5%)
High	219 (100%)	106 (79%)	69 (38%)	121 (95%)
Not Available	3	1	1	0

†There is an overlap of 59 neuroblastoma patients with both whole exome and whole genome sequencing. Patients with targeted sequencing are all unique and do not overlap with whole exome or whole genome cases, yielding 482 unique patients with exonic DNA sequencing of *LMO1*.
‡International Neuroblastoma Staging System (INSS).

Table 2.3: Replication and meta-analysis of rs2168101 association.

SNP	Ref/Alt	Cohort	MAF Cases	MAF Controls	Add *P*-value	Add Odds Ratio (T vs. G)	Het Odds Ratio (GT vs. GG)	Hom Odds Ratio (TT vs. GG)
rs2168101	G/T	Eu-Am†	0.242 (n=2101)	0.313 (n=4202)	4.14×10^{-16}	0.67 (0.61-0.74)	0.69 (0.62-0.77)	0.52 (0.42-0.64)
		Italian	0.164 (n=420)	0.250 (n=751)	2.07×10^{-06}	0.61 (0.50-0.75)	0.57 (0.44-0.74)	0.40 (0.21-0.75)
		U.K.	0.190 (n=369)	0.311 (n=1109)	5.86×10^{-10}	0.56 (0.47-0.68)	0.51 (0.39-0.66)	0.31 (0.18-0.53)
		Af-Am†	0.0865 (n=364)	0.0891 (n=2491)	0.20	0.79 (0.56-1.13)	0.96 (0.71-1.30)	1.07 (0.38-3.04)
		Combined			7.47×10^{-29}	0.65 (0.60-0.70)	0.67 (0.61-0.73)	0.49 (0.41-0.59)

MAF = minor allele frequency; Add = additive model; Het = heterozygous; Hom = homozygous; Eu-Am = European-America; Af-Am = African American
†Imputed genotypes and correction for population stratification.

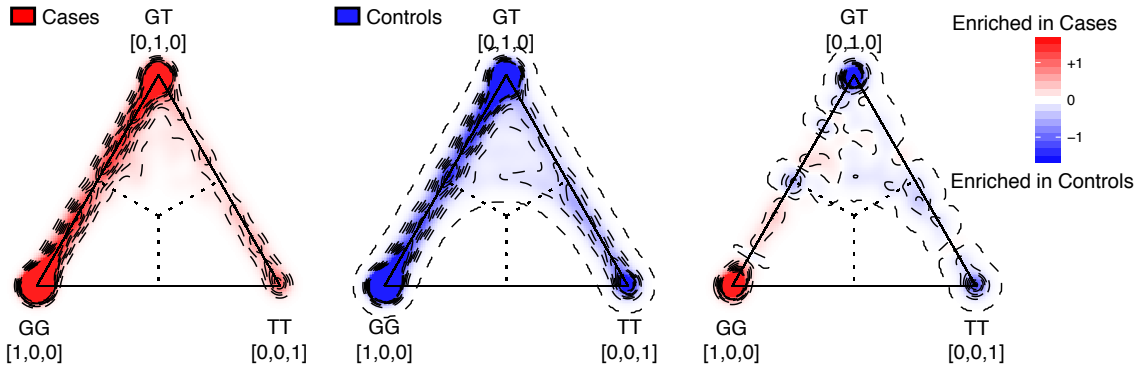

Figure 2.3: The imputed SNP, rs2168101, is associated with neuroblastoma, and the risk 'G' allele is enriched in neuroblastoma cases. Ternary density plots of genotype probability vectors [P(G/G), P(G/T), P(T/T)] output from IMPUTE2 for rs2168101 in the European-American cohort. Vertices represent 'perfect; confidence calls in which P(genotype) = 1; dotted lines represent decision boundaries for genotype calling based on most probable genotype. All plots were normalized by the total number of individuals studied and subjected to 2D Gaussian kernel smoothing. Left, 2,101 cases (red); centre, 4,202 controls (blue); right, difference between cases and controls highlights enrichment of G/G genotype (homozygous risk) in cases and of G/T and T/T genotypes in controls. Validation efforts using PCR-based genotyping in 146 out of 2,101 European-American cases confirmed an 86% concordance with imputation based on most probable genotypes.

evolutionary conservation, active enhancer localization, and disruption of a transcription factor binding motif, including the sentinel SNP rs110419 ($P = 1.17 \times 10^{-13}$) from our original report[50].

To test for the possibility of multiple statistical signals or enhancers not marked by conservation or p300 at the *LMO1* locus, we repeated association testing conditional on imputed rs2168101 genotypes and observed no significant variants after multiple test correction (most significant variant: rs34544683, nominal $P = 9.0 \times 10^{-4}$, Bonferroni $P = 1$; Figure 2.6a). To test whether the rs2168101 signal can be equally captured by other variants, we also performed reciprocal association tests for rs2168101 conditioned on all 27 other SNPs within 1.5 megabases (Mb) of *LMO1* passing thresholds MAF > 0.01 and nominal $P < 1 \times 10^{-5}$. Notably, rs2168101 remained significant across all

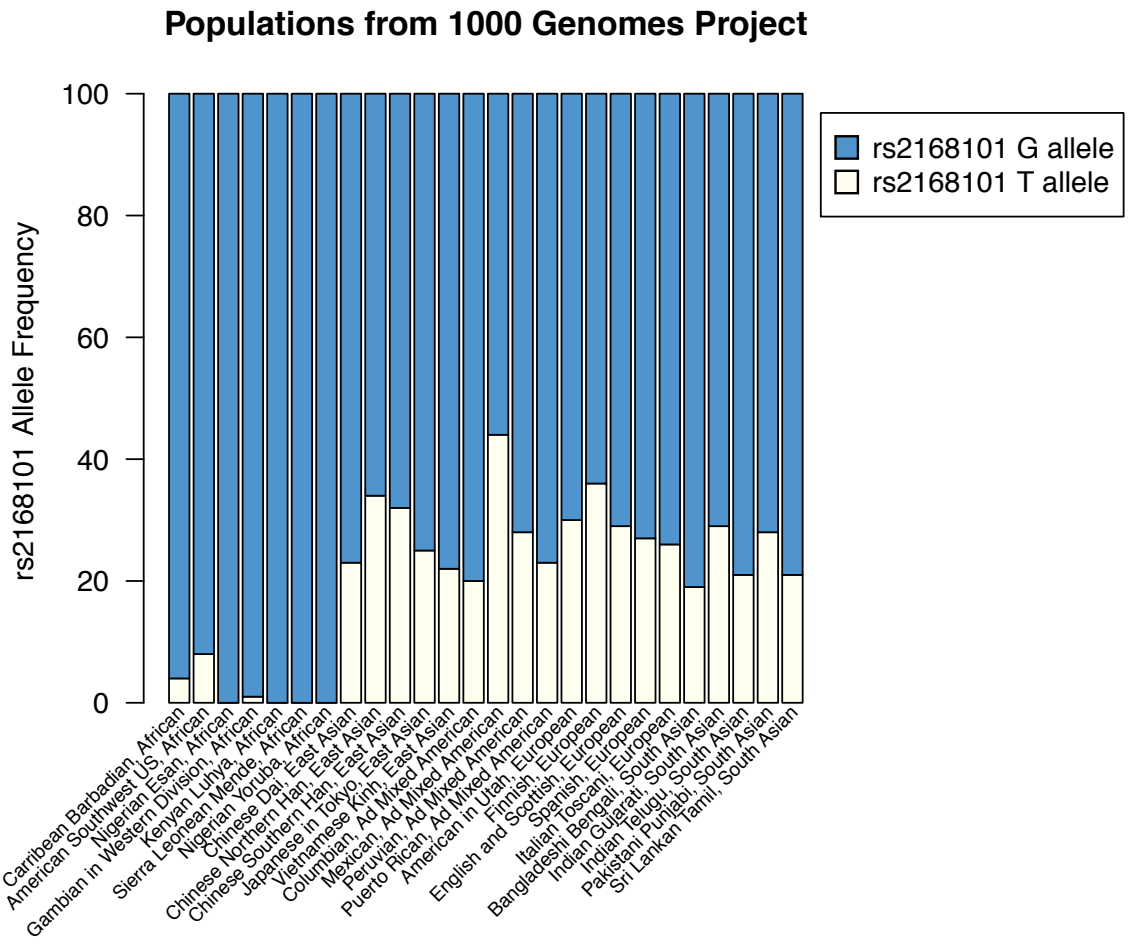

Figure 2.4: The protective 'T' allele of rs2168101 is rare or absent in African populations. Shown are allele frequencies for populations that were profiled as part of the 1000 Genomes Project. Comparative genomic analysis of other vertebrate species indicates that the risk G allele is ancestral, and the low frequency in African populations implies that the protective T allele is recently evolved in human history. `http://browser.1000genomes.org/Homo_sapiens/Variation/Population?db=core;r=11:8254908-8255908;v=rs2168101;vdb=variation;vf=1736493`

conditional tests (worst-case nominal $P = 2.6 \times 10^{-7}$, Bonferroni $P = 0.002$; Figure 2.6b). These results are consistent with a single underlying signal at the *LMO1* locus, and re-affirm that rs2168101 is the single best causal SNP candidate, because its association with neuroblastoma cannot be accounted for by other variants.

We next sought to determine whether rs2168101 genotypes were associated with *LMO1* expression by messenger RNA sequencing (mRNA-seq) of 127 primary high-risk neuroblastoma tumors. Genotyping rs2168101 yielded 102 G/G, 25 G/T and no T/T tumors (MAF = 9.8%). We observed significantly higher *LMO1* expression in G/G versus G/T genotype tumors (t-test $P = 0.028$; Figure 2.7a). Notably, the absence of protective homozygous T/T genotypes in this high-risk neuroblastoma cohort is consistent with our previous observation that the risk alleles predispose to the high-risk phenotypic subset[50] (for clinical covariate associations, see Table 2.4). Accordingly, the rs2168101 G/G genotype is highly associated with decreased neuroblastoma patient event-free ($P = 0.0004$) and overall ($P = 0.0004$) survival compared to G/T and T/T genotypes together in our European-American cohort (Figure 2.8). Two cell lines with homozygous T/T or T/- genotypes expressed *LMO1* at comparatively lower levels than cell lines containing the G allele (Figure 2.9a).

GATA transcription factors mediate chromatin looping and facilitate long-range enhancer-promoter interactions to regulate target gene expression[78]. We therefore sought to confirm allelic imbalance of *LMO1* transcripts (a hallmark of gene regulation *in cis*), which could result from differential GATA-binding caused by rs2168101. First, because the rs2168101 intronic SNP is not detectable by mRNA-seq, we identified the *LMO1* exonic synonymous SNP, rs3750952, which can measure allelic expression in the heterozygous state. We identified 45 tumors with the necessary rs3750952 = C/G genotype, and then directly genotyped rs2168101 (G/G = 33, G/T = 12, T/T = 0) in this panel. By mRNA-seq, there was greater allelic imbalance in 12 tumors that were heterozygous for rs2168101 (G/T) than in 33 homozygous tumors (rs2168101 = G/G; t-test $P < 0.0001$; Figure 2.7b). We next used targeted sequencing of nuclear-enriched nascent RNAs in four neuroblastoma cell lines (G/G = 1, G/T = 2, T/T = 1) to provide direct ascertainment of allele-specific expression at rs2168101.

Table 2.4: Association of rs2168101 with clinical/biological co-variates.

Clinical/Biological Co-variate	rs2168101 genotypes*			Association Result	
	GG	GT	TT	P-Value†	Odds Ratio†
Stage‡4	530 (62%)	280 (33%)	49 (6%)	0.01198	0.81 (0.69-0.95)
Not Stage 4	611 (56%)	400 (37%)	74 (7%)		
MYCN Amplified	183 (55%)	115 (34%)	36 (11%)	0.00297	1.39 (1.12-1.73)
MYCN Non-Amplified	881 (59%)	525 (35%)	83 (6%)		
High-Risk	523 (63%)	263 (32%)	47 (6%)	0.00174	0.76 (0.65-0.90)
Not High-Risk	594 (56%)	398 (37%)	73 (7%)		
Unfavorable Histology	454 (61%)	237 (32%)	48 (6%)	0.14479	0.88 (0.73-1.05)
Favorable Histology	527 (57%)	336 (36%)	62 (7%)		
DNA Index Hyperdiploid	685 (59%)	412 (35%)	71 (6%)	0.32009	0.91 (0.76-1.09)
DNA Index Diploid	324 (57%)	198 (35%)	43 (8%)		
Age $\geq$ 18 mos	621 (61%)	346 (34%)	55 (5%)	0.01448	0.82 (0.69-0.96)
Age $<$ 18 mos	529 (57%)	338 (36%)	68 (7%)		

*Reverse strand hg19, imputed genotypes from IMPUTE2, genotype frequencies as reported by SNPTEST.
†SNPTEST, frequentist score test with additive model, adjusted for gender and top 20 MDS components.
‡International Neuroblastoma Staging System (INSS).

In both heterozygous lines, we observed allelic imbalance that significantly favoured the risk G allele over the protective T allele (Figure 2.7c). Collectively, these results indicate that the intact GATA motif at rs2168101 results in significantly higher *LMO1* expression levels than the TATA coded by the alternative allele. Allelic imbalance of *LMO1* was not driven by somatic DNA alterations (for example, loss of heterozygosity) that could affect allelic dosage (Figure 2.9b).

Examination of neuroblastoma transcriptome data for 127 primary tumors showed that *GATA2* and *GATA3* are overexpressed compared to other members of the GATA transcription factor family (Figure 2.10a), and that *GATA3* is the most highly expressed. Additionally, protein immunoblotting showed that GATA3 is uniformly highly expressed in neuroblastoma cell lines, while LMO1 is highly expressed in the G/G

(SKNSH and SHSY5Y), G/- (KELLY) and G/T (IMR32) cell lines, but only barely detectable in the BE2C cell line that lacks a G allele at the rs2168101 locus (Figure 2.10b). We therefore performed ChIP-seq using a GATA3 antibody in neuroblastoma cell lines, and observed robust GATA3 binding at rs2168101 in lines containing the G allele (SHSY5Y, KELLY, BE2 and NGP) but not in a line containing only a T allele (BE2C; Figure 2.11a). We then specifically considered GATA3 ChIP-seq reads overlapping rs2168101, and we observed strong preferential binding to the G allele in the G/T heterozygous cell lines BE2 (0.97 G-allele fraction from 38 reads, 95% confidence interval: 0.86-1.00, Binomial test $P = 2.8 \times 10^{-10}$) and NGP (1.00 G-allele fraction from 6 reads, 95% confidence interval: 0.54-1.00, Binomial test $P = 0.03$; Figure 2.11b).

Acetylation of histone H3 at lysine 27 (H3K27ac) is a hallmark of active enhancers[79], and ChIP-seq analysis of SHSY5Y (G/G; not MYCN amplified), KELLY (G/-; MYCN amplified), BE2 (G/T; MYCN amplified) and NGP (G/T; MYCN amplified) neuroblastoma cells showed extensive H3K27 acetylation in the first intron of *LMO1* across rs2168101, which was not observed in BE2C (T/-; MYCN amplified; Figure 2.12a). This region is classified as a super-enhancer in G-allele-containing lines SHSY5Y, KELLY and BE2 based on enhancer clustering and especially high H3K27ac signal, a pattern also observed for other known oncogenes and tumor suppressor genes in this disease[80] (Figure 2.12b and Figure 2.13a). No super-enhancer was observed in BE2C, Jurkat T-ALL cells that also express LMO1[81], or in other non-neuroblastoma tissues from ENCODE (Figure 2.12b and Figures 2.13b,c). These results are consistent with recent evidence that disease-associated SNPs frequently affect enhancers that are specific to disease-relevant cell lines and tumour histology, and control developmental stage and tissue-specific gene expression[80, 82–86].

We next performed luciferase reporter assays to measure the effect of rs2168101

alleles on enhancer activity. HEK293T cells transfected with constructs containing the risk G allele demonstrated 30-300-fold higher normalized luminescence compared to the T allele (t-test $P = 0.002$, Figure 2.14a), whereas luciferase activity of the T allele was not significantly different from empty vector, indicating that the intact GATA motif is required for robust enhancer activity. Finally, knockdown of GATA3 in SHSY5Y and KELLY cells resulted in both decreased LMO1 protein levels and suppression of cell growth that was rescued by LMO1 overexpression (Figure 2.14b), indicating the central role of GATA3 in regulating LMO1 expression levels in neuroblastoma.

2.3 Summary and future directions

2.3.1 Clinical significance and avenues for translational research

Taken together, these data demonstrate the underlying molecular mechanism for a highly robust genetic association to neuroblastoma, mediated by a single common causal SNP rs2168101 that disrupts a GATA transcription factor binding site within a tissue-specific super-enhancer element. The risk allele is associated with a greater prevalence of high risk disease, and is associated with a worse prognosis independent of *MYCN*-amplification. Additionally, the rarity or absence of the protective allele in African populations and its relative depletion in African-Americans may partially explain the more aggressive clinical course in African-American children[87]. Moreover, this work further confirms the utility of association studies to define clinically relevant oncogenic pathways.

Transcriptional factors and co-regulators, such as *LMO1*, have not traditionally been considered "druggable" targets, because they are not enzymatic proteins and therefore

lack a catalytic domain to which small molecule inhibitors can be designed. Thus, it will be important to see if ongoing efforts to map the downstream targets of LMO1—for example, by profiling and integrating whole transcriptome changes that accompany induced LMO1 up- or down-regulation with LMO1 ChIP-seq to identify genes which are directly bound by LMO1-nucleated transcription factor complexes—may uncover druggable effectors of LMO1 signaling and suggest new therapeutic strategies. These LMO1 binding studies are already underway, as part of a newly R01-funded effort to investigate the broader regulatory landscape of LMO1 in neuroblastoma, both upstream and downstream.

Early reports that BET bromodomain inhibition can be therapeutically exploited in neuroblastoma in addition to other c-MYC- or MYCN-driven cancers provides another possible avenue for translational research. Indeed, the dependence of neuroblastoma cells on super-enhancer-mediated *LMO1* expression provides another potential mechanism for the sensitivity of these tumors to inhibitors of the transcriptional machinery such as CDK7 and BET bromodomain proteins[82, 84]. Moving forward, it will be interesting to investigate to what extent chromatin modulators may impinge on the GATA3-LMO1 enhancer and signaling axis, and may therefore provide another therapeutic avenue.

2.3.2 *In vivo* models for tumor initiation

The use of *in vivo* models to validate the effect of causal variants on cancer initiation has been limited by many technical challenges, including genetic, epigenetic, and physiological differences between human and possible model organisms; the relatively short duration of models compared to human tumorigenesis; difficulty in modeling complex interactions between multiple pathogenic variants; and the relatively low

penetrance/effect size of GWAS variants[88]. In this context, we have been fortunate to maintain a very productive collaboration with the laboratory of Thomas Look at Dana-Farber, where we have developed transgenic zebrafish models of neuroblastoma to investigate the role of known or candidate neuroblastoma oncogenes in the developing nervous system[89].

Such a zebrafish model was already successfully designed to investigate how mutated ALK and hyperactive MYCN cooperate to promote tumorigenesis[90], and in another model, constitutive co-overexpression of LMO1 and MYCN in the developing zebrafish decreased tumor latency and increased tumor penetrance relative to MYCN overexpression alone (Zhu *et al.*, submitted). The observation of metastasis in the LMO1/MYCN co-overexpressing zebrafish (Zhu *et al.*, submitted) but not in MYCN overexpressing zebrafish controls raises the tantalizing possibility that LMO1 may play important roles in promoting metastasis in human neuroblastoma as well, which remains to be confirmed or ruled out by future studies. In addition, our collaborators are actively working on zebrafish models of *LMO1* enhancer knockout, as well as introduction of the GATA-ablating T allele of rs2168101, and we are hopeful that these can be used to further validate the results of the present study.

2.3.3 The role of GATA transcription factors in neuroblastoma

GATA3 has attracted interest in neuroblastoma in recent years, both as a direct positive regulator of Cyclin D1(CCND1)[91] and for inhibiting differentiation and promoting stemness[92, 93]. The identification here of GATA3 as a positive regulator of LMO1 signaling elucidates another facet into how both of these genes function as oncogenes in neuroblastoma. Although our rescue experiments suggest that LMO1

may be the primary downstream regulator of GATA3-driven cell proliferation in neuroblastoma with an active *LMO1* enhancer, further studies will be necessary to clarify whether proposed GATA3-mediated effects on cellular differentiation operate in concert or separately from LMO1 and/or if GATA3 may operate differently in the absence of a functioning *LMO1* enhancer. In this regard, a crucial starting point will be to perform co-IP to determine if LMO1 and GATA transcription factors directly interact as part of the same protein complexes—as has been observed in the context of blood stem/progenitor cells for the better-studied LMO2[94]—and to parse out their combinatorial relationship in neuroblastoma by comparison of LMO1 and GATA2/GATA3 binding profiles.

Recently, a gapped k-mer SVM method has been developed to build robust models of how underlying DNA sequences can predict chromatin features[95, 96]. Important applications of this method include prediction of which transcription factors (inferred by their motifs) are active in a specific cell type, and whether such transcription factors tend to function predominantly as transcriptional activators or transcriptional repressors. We have now successfully applied this method to our neuroblastoma H3K27ac ChIP-seq data, which predicted that GATA transcription factors are one of the strongest transcriptional activators in neuroblastoma (see Figures 2.15 and 2.16). This preliminary analysis underscores the potential importance of GATA transcription factors in maintaining neuroblastoma transcriptional programs, likely at a global genome-wide scale beyond what we have observed at the *LMO1* locus, warranting further study.

2.3.4 The non-coding genome of neuroblastoma

The data and approach that were used here to identify the causal variant at the *LMO1* locus has helped lay the groundwork for the discovery of other functional germline variants and somatic driver mutations that affect non-coding regulatory regions of the genome in neuroblastoma. One method we have begun to explore for this purpose is deltaSVM[97], an extension of the gapped k-mer SVM method mentioned above that uses SVM weights to compute a score to predict whether an input mutation is likely to be activating (positive score), inactivating (negative score), or neutral (near zero score) when trained on either H3K27ac ChIP-seq or DNase-seq or ATAC-seq data. As a validation of this method, we used H3K27ac ChIP-seq data generated from 8 neuroblastoma cell lines to compute deltaSVM scores for the 27 top associated variants at the *LMO1* locus, which predicted the rs2168101 protective T allele to be strongly inactivating (Table 2.1 and Figure 2.17).

While the gapped k-mer SVM method (and deltaSVM, by extension) has good sensitivity and specificity, it suffers from low positive predictive value due to the overwhelming predominance of functionally inert DNA in the human genome[95]. One way in which this problem can be addressed is to restrict the deltaSVM analysis only to non-coding regions with a high likelihood of being functionally active, such as open chromatin regions. To this end, we are pursuing ATAC-seq experiments—a new transposase-based methodology which combines the advantages of DNase-seq (assessing chromatin accessibiliy) and MNase-seq (assessing nucleosome positioning)[98]—to map the chromatin architecture of neuroblastoma with higher resolution than is possible with current H3K27ac ChIP-seq data.

We are also working to integrate other functional data as additional supportive evidence of functional non-coding genetic variants or mutations, including expression

data from TARGET. One general approach is to look for expression outliers that are associated with specific genetic changes, which can indicate a potential regulatory phenotype. While the low recurrence of non-coding mutations (similar to coding mutations) in neuroblastoma presents a difficult challenge, we have already identified a few possible leads, including a TARGET patient with a mutation in the proximal promoter of *MALAT1* that is also a *MALAT1* expression outlier (Figure 2.18); MALAT1 stands for "metastasis associated lung adenocarcinoma transcript 1" and is a long non-coding RNA that has already been extensively associated with cancer[99, 100]. A few additional patients also appear to be expression outliers by RNA-seq but are unfortunately missing paired WGS data (Figure 2.18), and they will therefore require targeted sequencing to discern whether their outlier status may also be associated with promoter mutations. We are also working to integrate allelic imbalance to specifically look for *cis*-regulatory signatures from RNA-seq data, though in practice this is not always possible for any given patient or gene because it depends on the presence of expressed heterozygous SNPs.

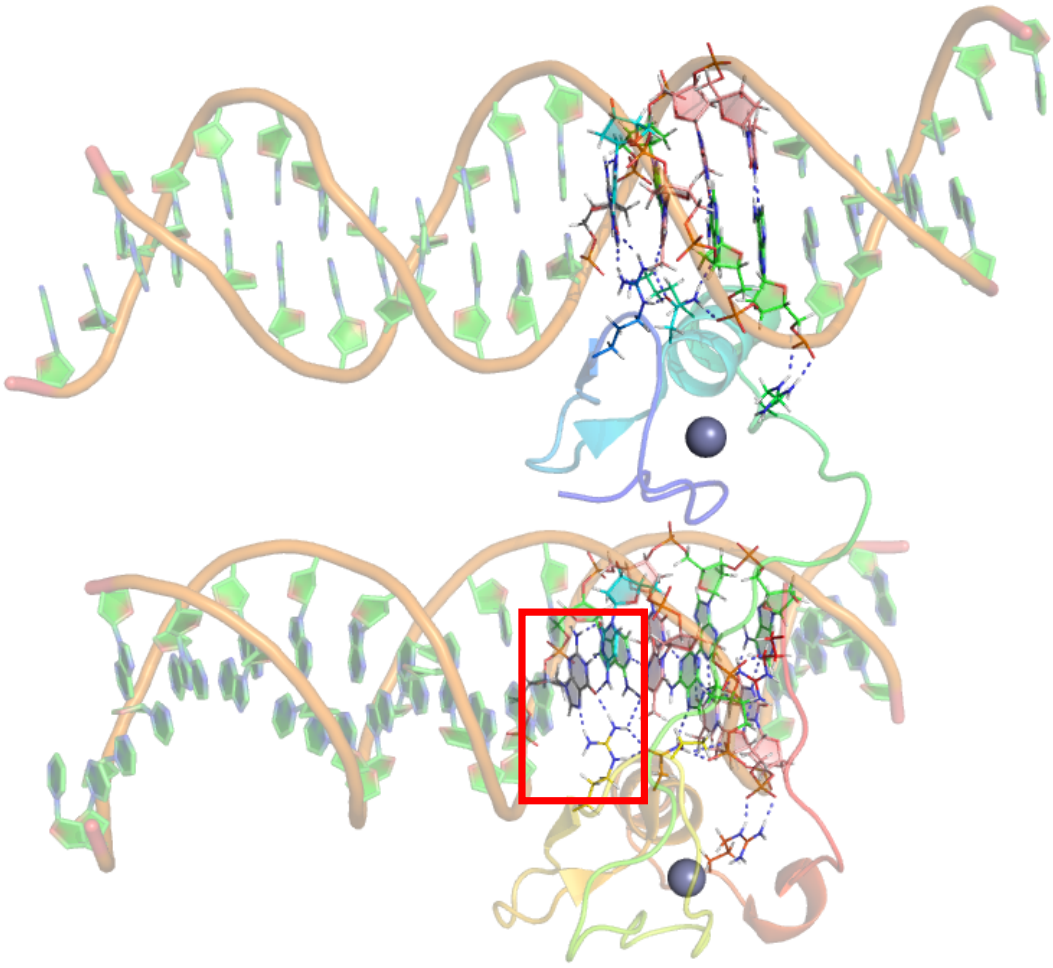

Figure 2.5: Co-crystal structure of GATA3 bound to its cognate 5'-AGATAA-3' DNA binding motif. GATA3 is drawn as a ribbon, illustrating the N-terminal zinc finger (blue) and C-terminal zinc finger (red) binding to two separate DNA molecules, which may reflect how GATA3 is able to facilitate long-range chromatin looping[75]. The red box highlights where the C-terminal zinc finger arginine 330 residue makes contact with the 5'-AGATAA-3' motif guanine, forming two hydrogen bonding contacts. An analogous contact occurs with arginine 276 of the N-terminal zinc finger. Interestingly, GATA3 R276P mutations are a cause of hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome[76], reinforcing the importance of this contact for maintaining proper GATA3 function. Image rendered in PyMOL from Protein Data Bank[77] structure: 4HC9.

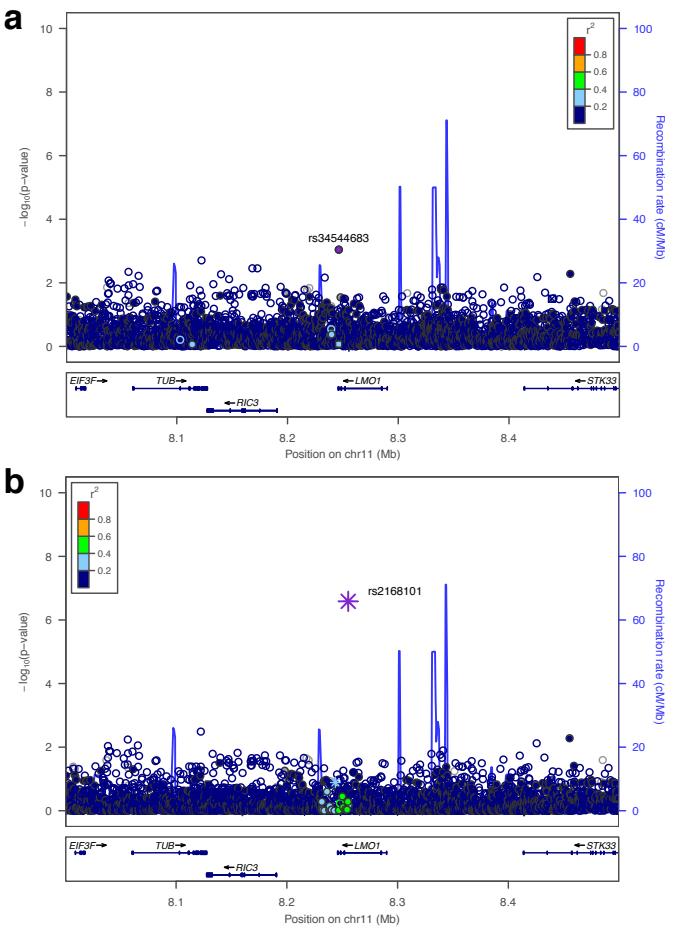

Figure 2.6: Conditional analysis reveals a single neuroblastoma association signal at the *LMO1* locus. a, Imputation-based neuroblastoma association study conditional on rs2168101. No variants remain significant after conditioning on rs2168101 (most significant variant: rs34544683, nominal $P = 9.0 \times 10^{-4}$, Bonferroni $P = 1$). **b**, Reciprocal analysis conditioned on each of 27 SNPs with a nominal $P < 1 \times 10^{-5}$. For rs2168101, the maximum (least significant) P-value across all non-rs2168101 conditional tests is shown, illustrating the extent to which the rs2168101 signal can be accounted for by other variants (a similar statistic is plotted for other variants). Notably, rs2168101 remained significant (worst-case nominal $P = 2.6 \times 10^{-7}$, Bonferroni $P = 0.002$) across all tests. These results are consistent with a single underlying signal at the *LMO1* locus, and re-affirm that rs2168101 is the single best causal SNP candidate because its association with neuroblastoma cannot be accounted for by other single variants.

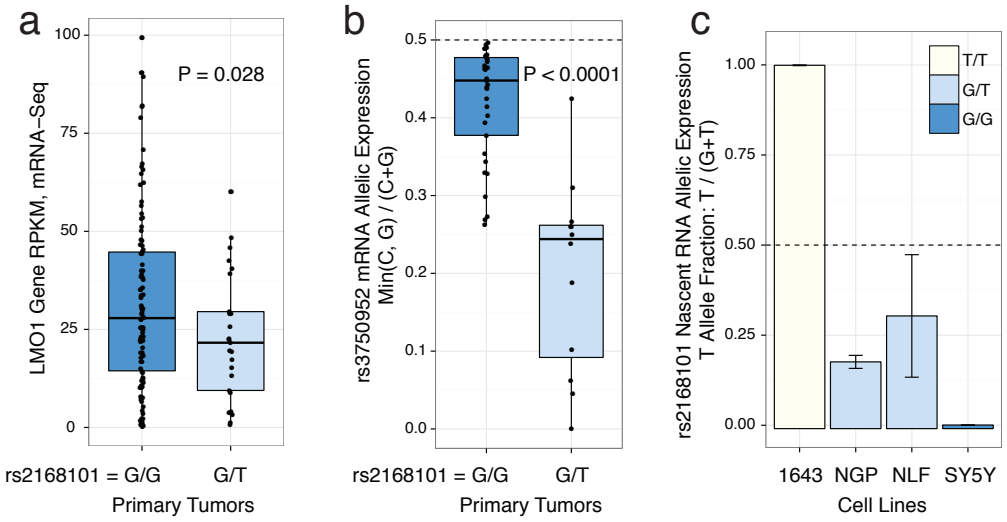

Figure 2.7: Allele-specific expression analysis confirms *cis*-regulatory signature at the *LMO1* locus. a, mRNA-seq across 127 primary tumours genotyped for rs2168101 (G/G = 102, G/T = 25, T = 0) revealed a significant decrease in LMO1 gene expression between G/T and G/G tumours (t-test $P = 0.028$). RPKM, reads per kilobase per million reads. **b**, Using the synonymous exonic SNP, rs3750952, to measure allelic expression by mRNA-seq revealed significantly more allelic imbalance in 12 heterozygous neuroblastoma tumours (rs2168101 = G/T) than in 33 homozygous tumours (rs2168101 = G/G) (t-test $P = 5.3 \times 10^{-5}$). **c**, Allelic expression for rs2168101 from targeted nascent RNA-seq in four neuroblastoma cell lines. The two heterozygous cell lines (rs2168101 = G/T) exhibited significantly reduced T-allele expression compared to the G allele (t-test $P = 1.6 \times 10^{-4}$ and 1.5×10^{-2} for NGP and NLF, respectively; error bars denote 95% confidence intervals across $n = 3$ duplicate experiments).

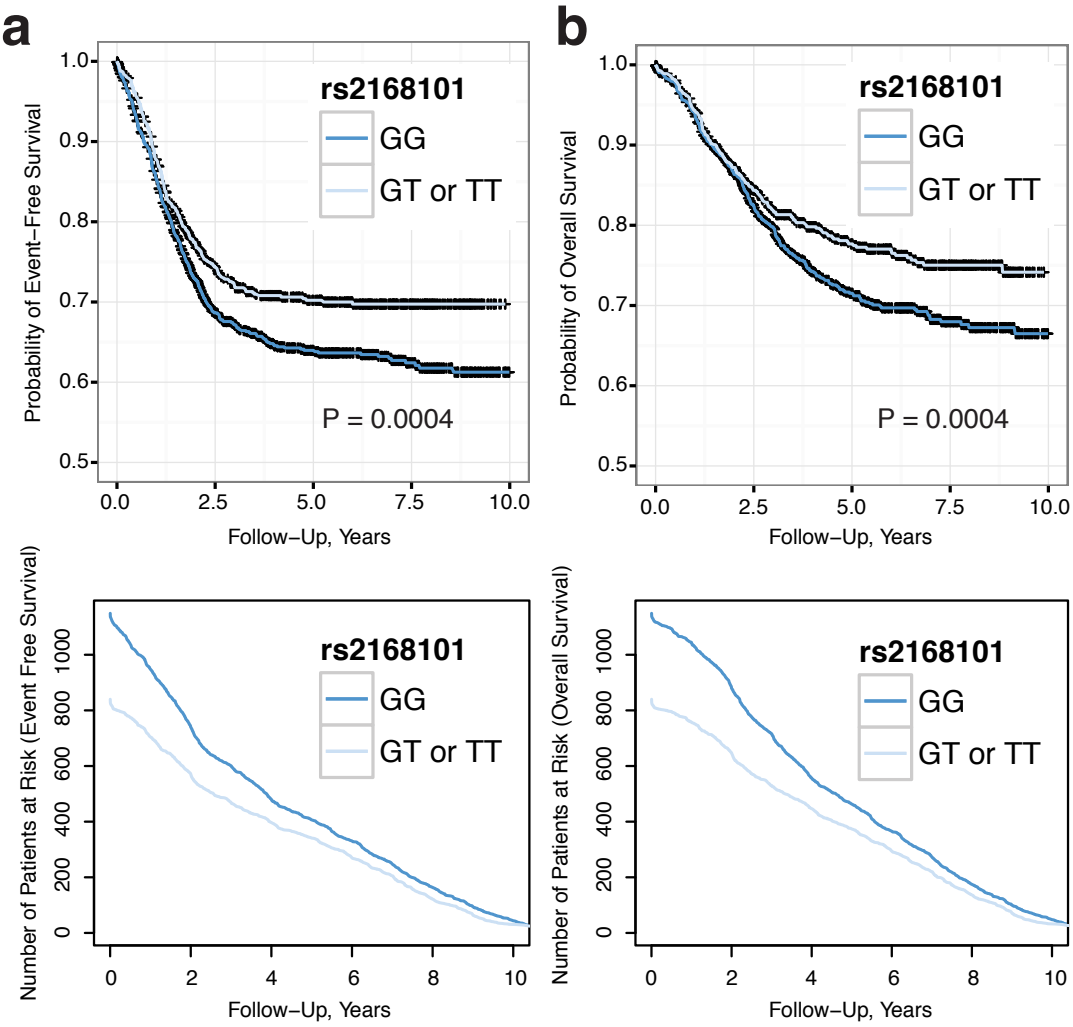

Figure 2.8: The protective T allele of rs2168101 is associated with increased event-free and overall survival in the European-American discovery cohort. Because genotypes for rs2168101 are imputed within the European-American discovery cohort, the most likely genotype for each neuroblastoma case was called based on the maximum of P(G/G), P(G/T) and P(T/T) from IMPUTE2. *P*-values reflect Cox proportional hazards regressions adjusted for *MYCN* amplification status and the first 20 MDS components to adjust for population stratification. **a**, Kaplan-Meier plot for event-free survival. Neuroblastoma cases with rs2168101 = G/G versus rs2168101 = G/T or T/T showed significantly worse event-free survival ($P = 0.0004$). **b**, Kaplan-Meier plot for overall survival. Neuroblastoma cases with rs2168101 = G/G versus rs2168101 = G/T or T/T showed significantly worse overall survival ($P = 0.0004$). Censored data points are shown as black crosses. Number of at risk patients at every time point for both event-free survival and overall survival are plotted below each respective Kaplan-Meier plot.

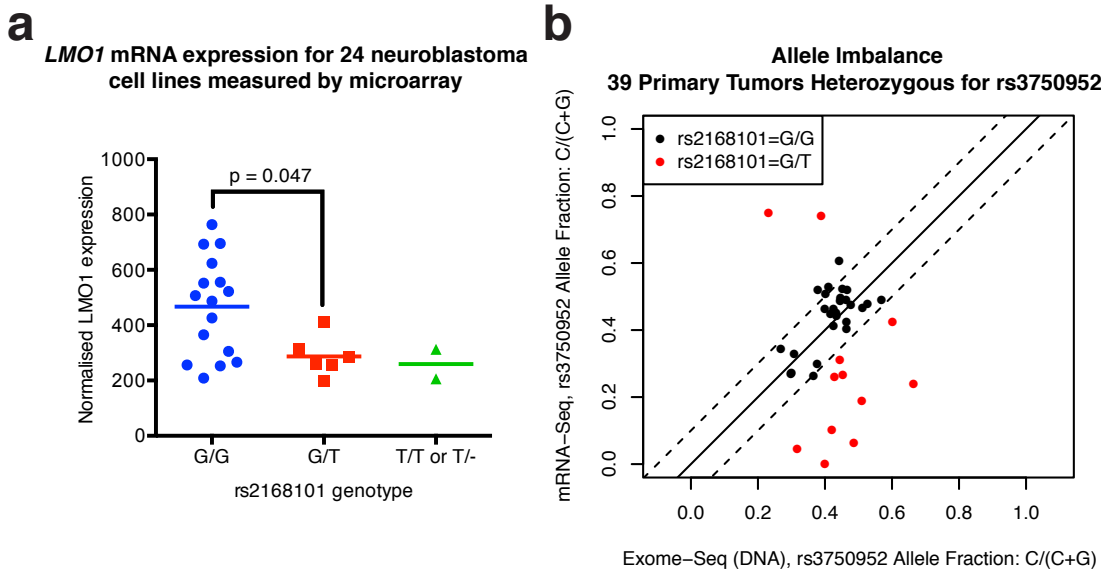

Figure 2.9: rs2168101 genotype is associated with total and allele-specific *LMO1* expression in neuroblastoma cell lines and primary tumours, and allele-specific expression differences are not driven by somatic DNA copy number alterations. a, Neuroblastoma cell line *LMO1* mRNA expression as quantified by Affymetrix U95Av2 oligonucleotide arrays was significantly higher in cell lines harbouring homozygous risk alleles (G/G) compared to heterozygous alleles (G/T) ($P = 0.047$, Mann-Whitney two-tailed). **b**, Allele-specific expression measured by RNA-seq from primary neuroblastoma tumours. Since rs2168101 is an intronic SNP that is spliced out in mRNA, the synonymous exonic SNP rs3750952 was used as a surrogate for measuring allele-specific expression in 39 primary tumours which are heterozygous for rs3750952 (C/G genotype). The DNA allelic fraction for rs3750952 determined by whole exome sequencing is plotted on the x-axis, whereas the RNA allele fraction for rs3750952 determined by mRNA-seq is plotted on the y-axis. The solid line indicates where DNA and RNA allele fractions are equal and dotted lines indicate the boundary where DNA and RNA allele fractions are within 10% of each other. Tumors that are heterozygous for rs2168101 (G/T genotype, red dots) exhibit greater RNA allelic imbalance ($P = 5.3 \times 10^{-5}$) than homozygous controls (rs2168101 = G/G genotype, black dots). By contrast, DNA allelic imbalance is no different between G/T versus G/G tumours ($P = 0.79$), indicating that a *cis*-acting regulatory mechanism, rather than somatic DNA alterations, drives *LMO1* allelic expression differences.

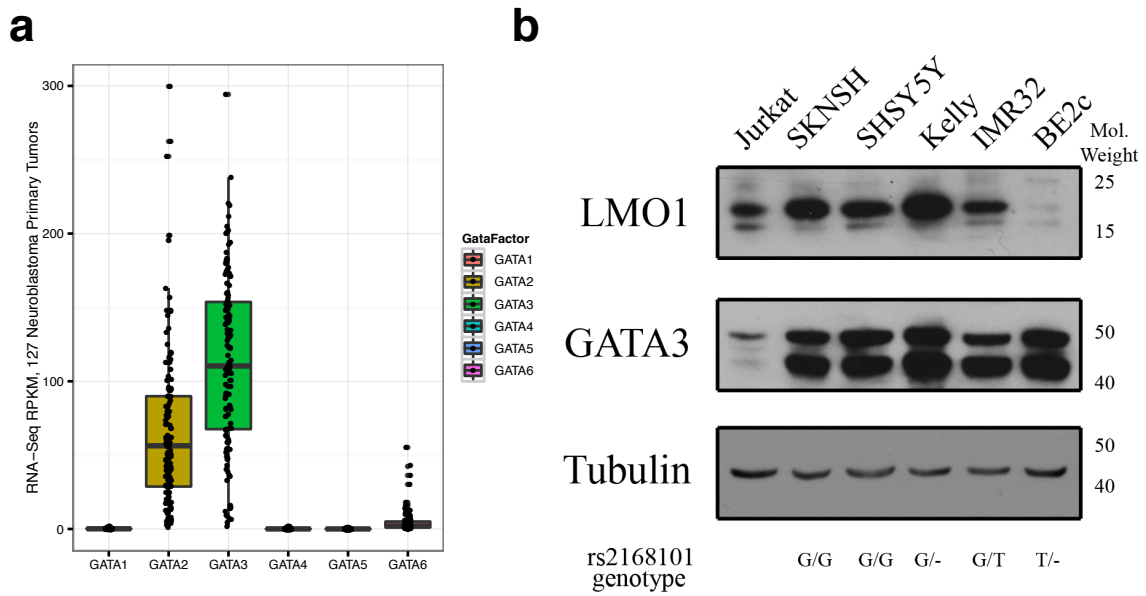

Figure 2.10: Expression of LMO1 and GATA-family transcription factors in neuroblastoma primary tumours and cell lines. a, RPKM expression measurements from mRNA-seq are summarized via boxplots for 127 primary neuroblastoma tumours for paralogues *GATA1* through *GATA6*. Both *GATA2* (median RPKM: 56) and *GATA3* (median RPKM: 110) are more highly expressed by 1-4 orders of magnitude on average compared to other members of the GATA family in neuroblastoma. **b**, Neuroblastoma cell lines were lysed for protein and resolved by SDS-PAGE. Jurkat T-ALL cells are shown as a positive control for LMO1 and GATA3 expression. Data are representative of at least three independent blots. The rs2168101 genotype is shown below individual cell lines.

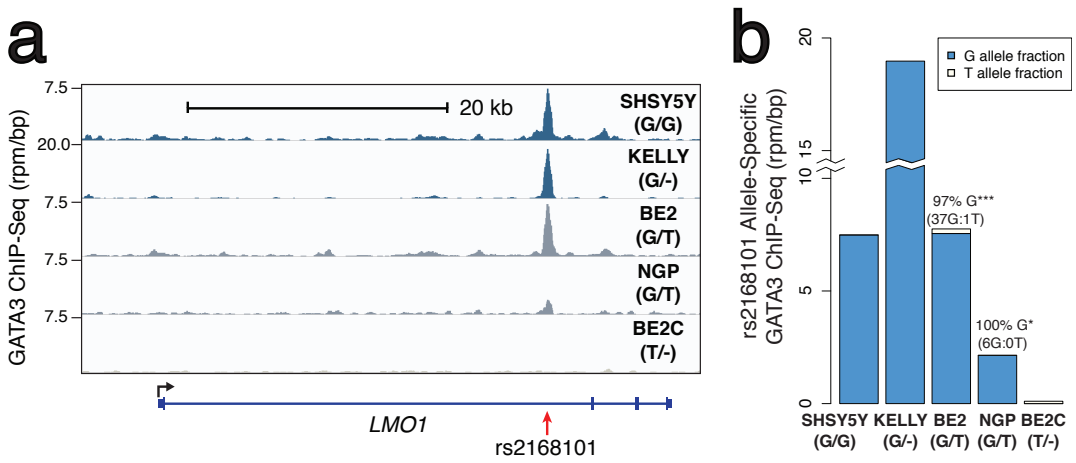

Figure 2.11: The rs2168101 protective T-allele negatively associates with GATA3-binding. a, Normalized GATA3 ChIP-seq signal at rs2168101 in five neuroblastoma cell lines (rs2168101 genotypes: SHSY5Y = G/G, KELLY = G/-, BE2 = G/T, NGP = G/T, BE2C = T/-). rpm, reads per million. **b**, Allele-specific binding of GATA3 at rs2168101. GATA3 binding highly favored the risk G allele in heterozygous lines (BE2: 0.97 G-allele fraction from 38 reads, 95% confidence interval: 0.86-1.00, Binomial $P = 2.8 \times 10^{-10}$; NGP: 1.00 G-allele fraction from 6 reads, 95% confidence interval: 0.54-1.00, Binomial $P = 0.03$).

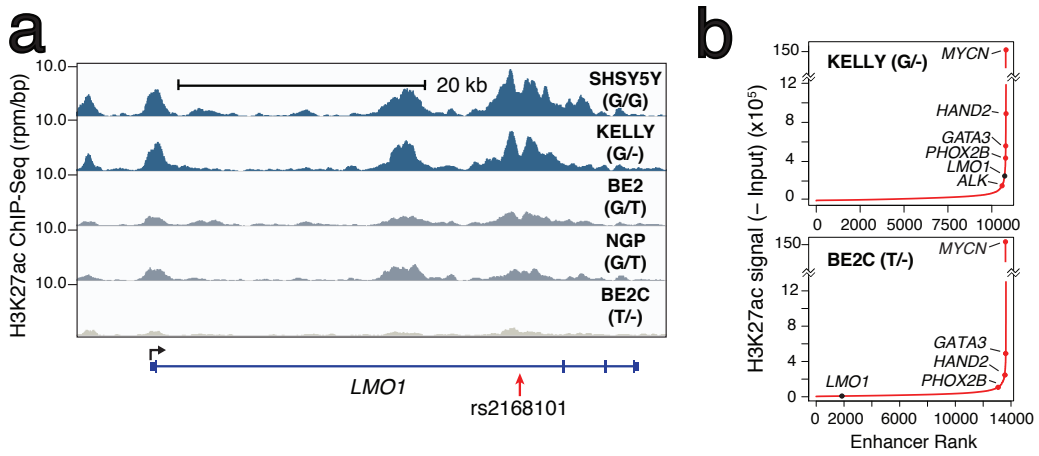

Figure 2.12: The rs2168101 protective T-allele negatively associates with the *LMO1* super-enhancer in neuroblastoma cells. a, Normalized ChIP-seq signal for H3K27ac at rs2168101. **b**, Ranked H3K27ac signal across all enhancers in MYCN-amplified KELLY and BE2C lines. Super-enhancers associate with key neuroblastoma genes, highlighted on the curve. There is an *LMO1*-associated super-enhancer in G-allele-containing lines SHSY5Y, KELLY and BE2, but not in BE2C, which lacks the G allele.

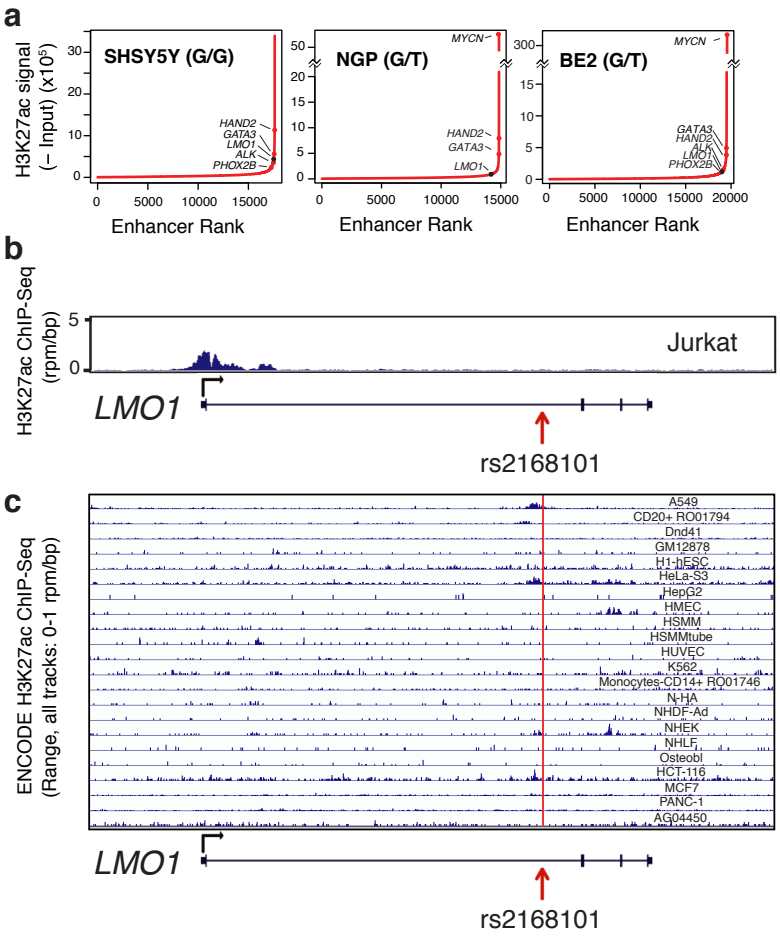

Figure 2.13: The *LMO1* super-enhancer is observed in neuroblastoma cell lines containing the G allele of rs2168101 and is highly tissue-specific. a, H3K27ac signal across all enhancers in SHSY5Y (*MYCN* not amplified; rs2168101 = G/G), BE2 (*MYCN*-amplified; rs2168101 = G/T) and NGP (*MYCN*-amplified; rs2168101 = G/T) is shown. Enhancers are ranked by their signal of H3K27ac minus input signal and are geometrically divided into two populations. Super-enhancers are those at the high end of the population and are associated with key genes in neuroblastoma, highlighted on the curve. LMO1-associated super-enhancers were identified in BE2, KELLY and SHSY5Y cells, which all contain the G allele of rs2168101, but not in BE2C cells in which the G allele is absent. **b**, H3K27ac ChIP-seq in the Jurkat cell line. **c**, All ENCODE non-neuroblastoma cell lines with H3K27ac ChIP-seq profiling. All non-neuroblastoma cell lines considered showed little to no evidence for an active enhancer element within the first intron of the *LMO1* gene locus, consistent with a tissue and disease-specific enhancer overlying the neuroblastoma causal SNP rs2168101.

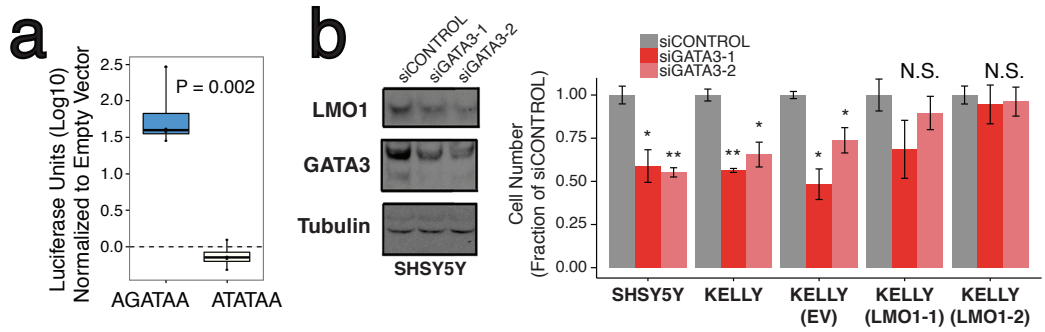

Figure 2.14: a, Luciferase reporter assay for LMO1 enhancer region. The risk G allele preserved enhancer activity (t-test $P = 0.002$ across $n = 4$ independent clones, each with $n = 5$ technical replicates), whereas the protective T allele was indistinguishable from empty vector. **b**, Left, protein blots for GATA3, LMO1 and tubulin in SHSY5Y cells treated with control (siControl) short interfering RNAs (siRNAs), or with siRNAs targeting GATA3 (siGATA3-1 and siGATA3-2), at 72 h post-treatment. Right, cell counts for cell lines SHSY5Y, KELLY, KELLY stably overexpressing control vector (EV) and KELLY with forced LMO1 overexpression (LMO1-1 and LMO1-2) treated with siRNAs at 72 h post-transfection. Rescue of suppressed cell growth after GATA3 depletion by forced LMO1 expression was observed at 72 h. Error bars denote ± s.e.m. $^{*}P < 0.05$, $^{**}P < 0.001$ by t-test. $n = 3$ independent transfections, $n = 9$ technical replicates.

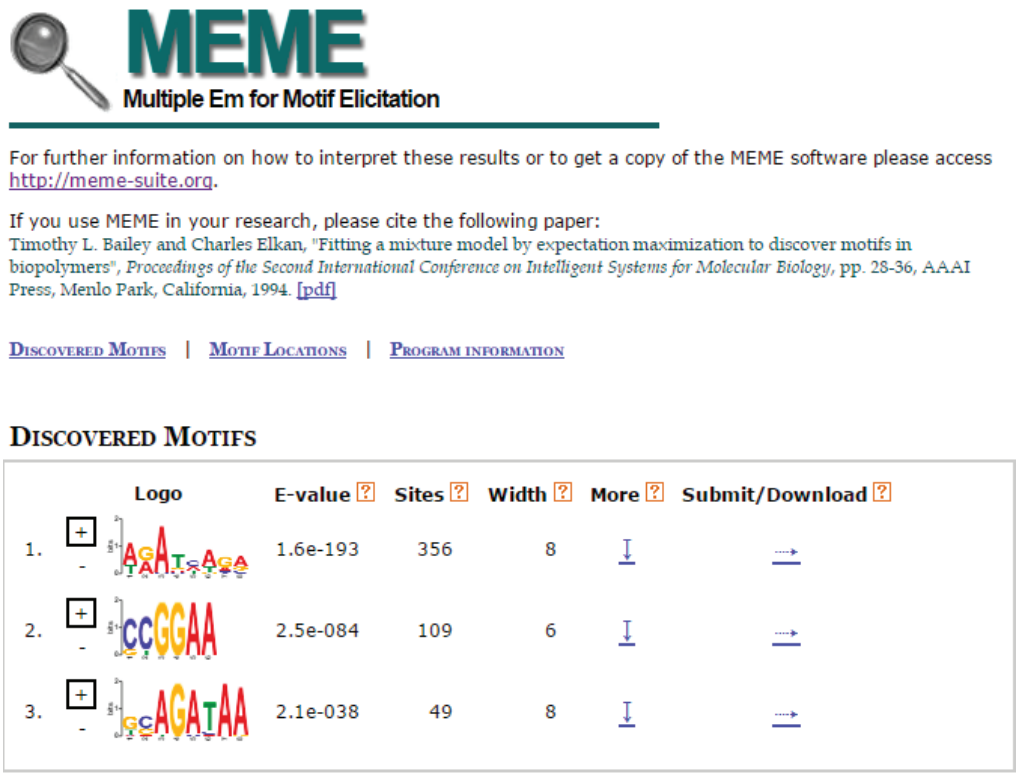

Figure 2.15: GATA transcription factor binding motifs are the sequences most highly associated with active chromatin regions in Kelly neuroblastoma cell line. These data reflect training gapped k-mer SVM on H3K27ac ChIP-seq data from the Kelly neuroblastoma cell line. Following SVM training, the top 1000 10-mer sequences with the strongest positive weights were input into MEME for unbiased motif discovery. Among the top three motifs, the first and third both match well to GATA transcription factors (5'-AGATAA-3'). The second motif is consistent with ETS transcription factors (5'-GGAA-3').

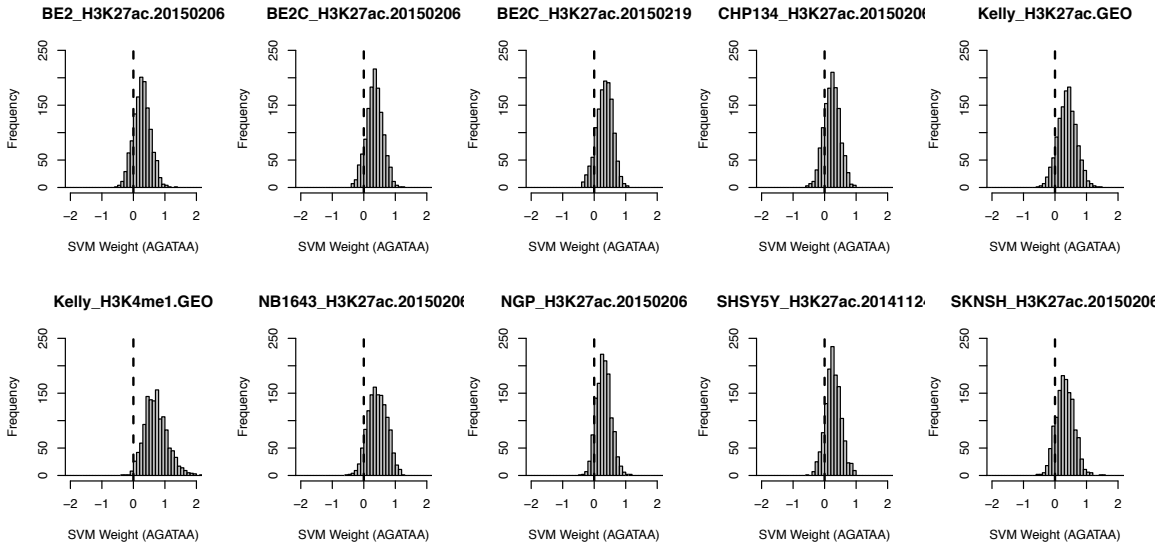

Figure 2.16: GATA transcription factor binding motifs are universally activating in neuroblastoma cell lines. These data reflect training gapped k-mer SVM on H3K27ac or H3K4me1 ChIP-seq data from neuroblastoma cell lines as indicated. Histograms reflect gapped k-mer SVM weights for all 10-mer sequences containing a 5'-AGATAA-3' motif or its reverse complement, 5'-TTATCT-3'. In all cell lines tested, the AGATAA motifs are biased toward higher weights, indicating that they are strongly associated with H3K4me1 and H3K27ac bound regions and suggesting that GATA transcription factors are likely to function as strong transcriptional activators at a genome-wide scale in neuroblastoma.

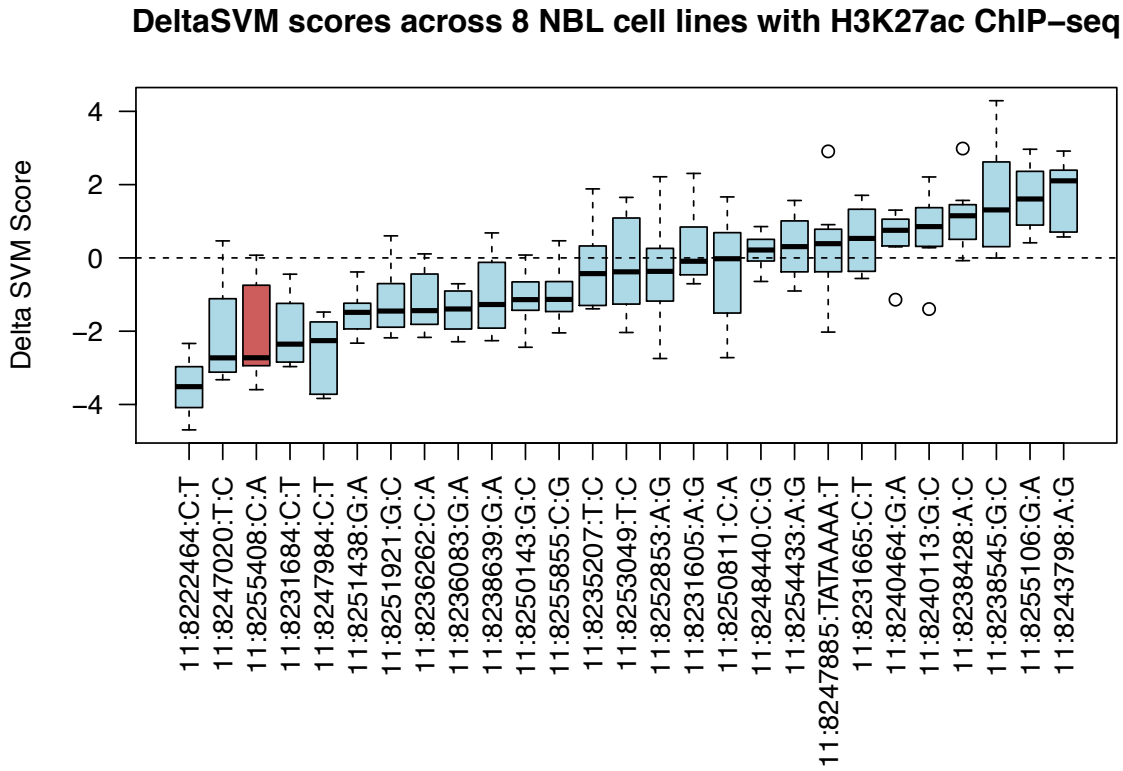

Figure 2.17: DeltaSVM trained on H3K27ac ChIP-seq data predicts that the protective T allele of rs2168101 is transcriptionally inactivating. DeltaSVM scores are shown by boxplot based on gapped k-mer SVMs trained on H3K27ac ChIP-seq profiles across 8 neuroblastoma cell lines for 27 candidate causal variants at the *LMO1* locus (Table 2.1). All 27 candidate causal variants are ranked by their median deltaSVM score across cell lines. The rs2168101 G→T SNP, predicted to be inactivating (strong negative score), is highlighted in red.

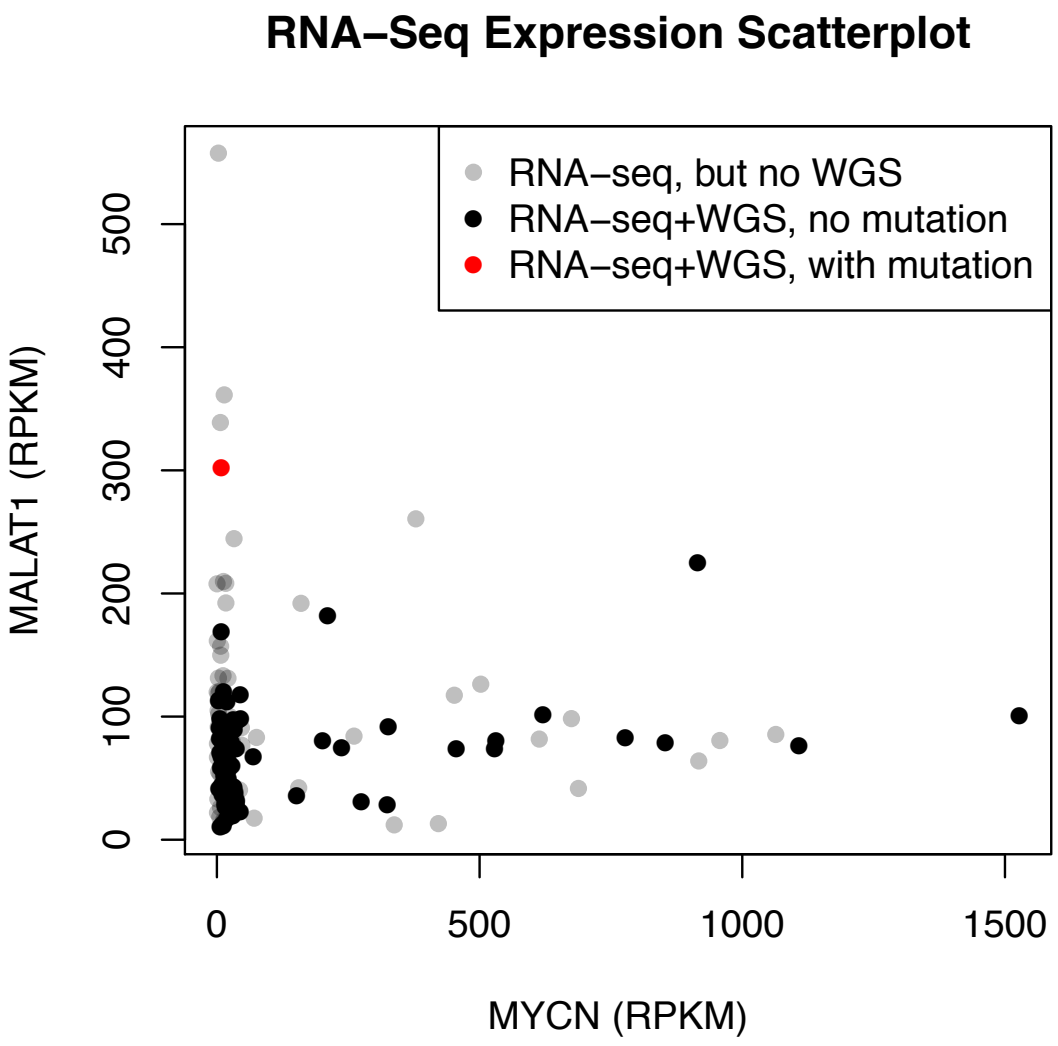

Figure 2.18: *MALAT1* expression outlier associated with proximal promoter mutation. Scatterplot reflects RPKM expression values as inferred by RNA-sequencing of TARGET primary tumors for *MALAT1* (y axis) versus *MYCN* (x axis). Several expression outliers are observed, and the highest *MALAT1* expressing TARGET tumor (patient PASPER) contains an insertion in its proximal promoter region that disrupts a TALE homeobox motif, which could potentially lead to overexpression of *MALAT1* by a removal of a repressive transcriptional regulator. Other expression outliers are identified, but unfortunately do not have paired WGS data (gray dots) in order to confirm whether or not they also contain mutations in the *MALAT1* proximal promoter.

Chapter 3

Clonal evolution in relapsed neuroblastomas.

3.1 Challenges for studying the relapsed neuroblastoma genome

Patients with high-risk neuroblastoma have a survival rate of less than 50%[1, 101], despite extensive treatment involving chemotherapy, surgery, radiation therapy and immunotherapy. In a majority of patients, an initial response to therapy is observed; however, up to 60% of these patients subsequently relapse with therapy-resistant tumors with very poor prognosis[7, 102, 103]. Recently, sequencing of the *ALK* locus in neuroblastomas at the time of relapse identified 14 activating mutations in 54 cases (26%)[104], suggesting that the frequency of *ALK* aberrations is higher in relapsed neuroblastoma genomes. However, nearly all next generation sequencing studies of neuroblastoma genetics to date have focused exclusively on primary tumors[44–47], due to the scarcity of relapsed neuroblastoma tissue both in the USA and worldwide. This reflects historical clinical thinking that biopsies are not indicated for relapsed neuroblastoma, given the prevalence of non-invasive diagnostic radiographic techniques[105], as well as the limited potential for a biopsy to change the clinical course given that

relapsed neuroblastoma is nearly always fatal[1]. Thus, the neuroblastoma research field has been faced with a long-standing dilemma: the development of new treatments for relapsed disease has been hindered by the absence of biopsy material, but a biopsy has not been ethically indicated unless it would possibly benefit the patient and affect treatment.

Despite the aforementioned rarity of relapse biopsies, our laboratory in collaboration with the TARGET project and the Children's Oncology Group (COG) were able to pull together matched normal blood, primary tumor, and relapsed tumor tissue "trios" from nine neuroblastoma patients. This enabled us to perform whole genome sequencing in order to define the germline and somatic lesions that define relapsed disease, and I took the lead on the data analysis for this project as part of my dissertation work. While presenting our results at the 2014 Advances in Neuroblastoma Research (ANR) Research Congress, we connected with colleagues in the Netherlands and France who had performed similar WGS analysis in six and eight trios, respectively. Our meeting ultimately led to a very productive collaboration between our three groups[106], which will be described in greater detail in the following sections.

3.2 Estimating the clonality/sub-clonality of mutations detected by next generation sequencing

In addition to performing more traditional mutation and copy number analyses in this dataset, I took the lead in a formal clonality analysis of somatic coding mutations across primary and relapse tumors. This provided an important opportunity to observe how neuroblastomas evolve molecularly over the course of therapy and to identify candidate mutations that may provide a selective advantage and ultimately lead to treatment resistance. The statistical methodologies that form the basis of this analysis

have only been developed in the past few years[107–109] and more details about these methodologies can be found in Appendix A. This section will focus on a illustrative and practical example of how these approaches have aided our understanding of clonal structure and evolution in neuroblastoma.

A useful metric for quantifying the relative clonality vs subclonality of mutations is the cancer cell fraction (CCF), which describes the fraction of the cancer cell population within a sample that harbor of a mutation. Thus, a clonal mutation is one with a CCF 100% (or estimated to be very close to 100%, practically) and a subclonal mutation is one with a CCF strictly less than 100%. Generally, this CCF quantity will be larger than the sample fraction of the mutant population, because a non-zero fraction of normal contaminating cells (containing only wild-type alleles by definition) will dilute the apparent prevalence of the mutation when counting mutant versus wild-type alleles from sequencing data. This problem of normal contamination, in addition to overall tumor ploidy, are the primary reasons why it is difficult to estimate mutation CCF from NGS data.

In practice, our laboratory has used Sequenza[109] in order to estimate tumor purity and ploidy from NGS data as a necessary first step, followed by statistical modeling laid out in Carter *et al.* and Landau *et al.* to estimate somatic point mutation CCF[107, 108]. The results of running Sequenza on the primary and relapse tumors of neuroblastoma patient "PASGAP" are shown in Figure 3.1. Although the B-allele frequency and relative coverage signals in the primary tumor data of PASGAP are diluted by somewhat lower sample purity (estimated purity of 43% in primary, 75% in relapse), the overall contamination-corrected profiles are largely concordant and confirm the expected clonal relationship between primary and relapse tumors.

A closer look at the PASGAP primary and relapse tumors reveals complex copy number alterations on chromosome 17 (see Figure 3.2). What is particularly interesting

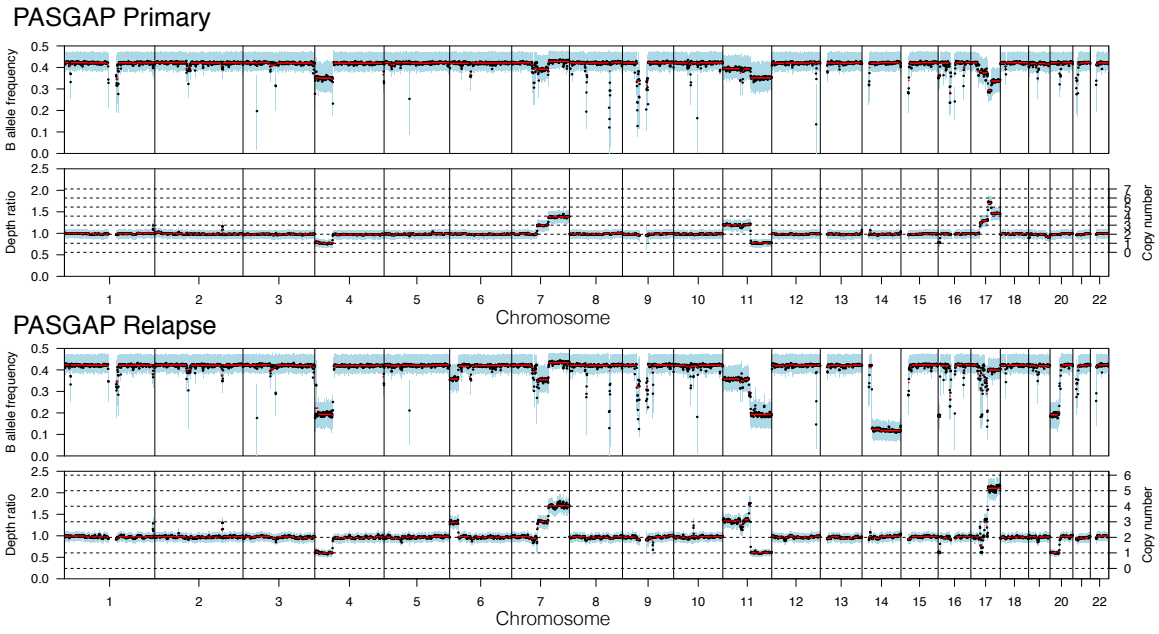

Figure 3.1: Sequenza results for the primary and corresponding relapse tumors from patient "PASGAP". Illustrated are genomic profiles of B-allele frequency of germline heterozygous SNPs (top) and relative sequencing coverage ratio (bottom) for primary and relapse tumors. Dotted lines illustrate Sequenza absolute copy number calls, which take into account normal contamination and overall tumor ploidy. These results illustrate that after taking into account the relatively lower purity of the primary tumor, copy number profiles are largely consistent between primary and relapse tumors, confirming their clonal relationship. Relapse-specific copy number alterations include gain of 6p, loss of 20p, and complex rearrangements across chromosome 17.

is that there is a region of relapse-specific hemizygous deletion (which is likely clonal, see Figure 3.2) and loss of heterozygosity that entirely covers the *NF1* tumor suppressor, which also contains a relapse-specific splice site mutation. While these data suggest that a sub-population of the PASGAP relapse tumor harbors bi-allelic loss of *NF1* function, it is unclear whether or not this *NF1* mutation is clonal, given that only 10 out of 15 reads in the relapse tumor show the mutation. However, after adjusting for tumor purity and ploidy, the CCF estimate with maximum posterior probability was indeed 100% (see Figure 3.3), indicating that this mutation is very likely clonal.

Thus, through this methodology we were able to say with some statistical confidence that patient PASGAP underwent clonal, bi-allelic inactivation of the *NF1* tumor suppressor.

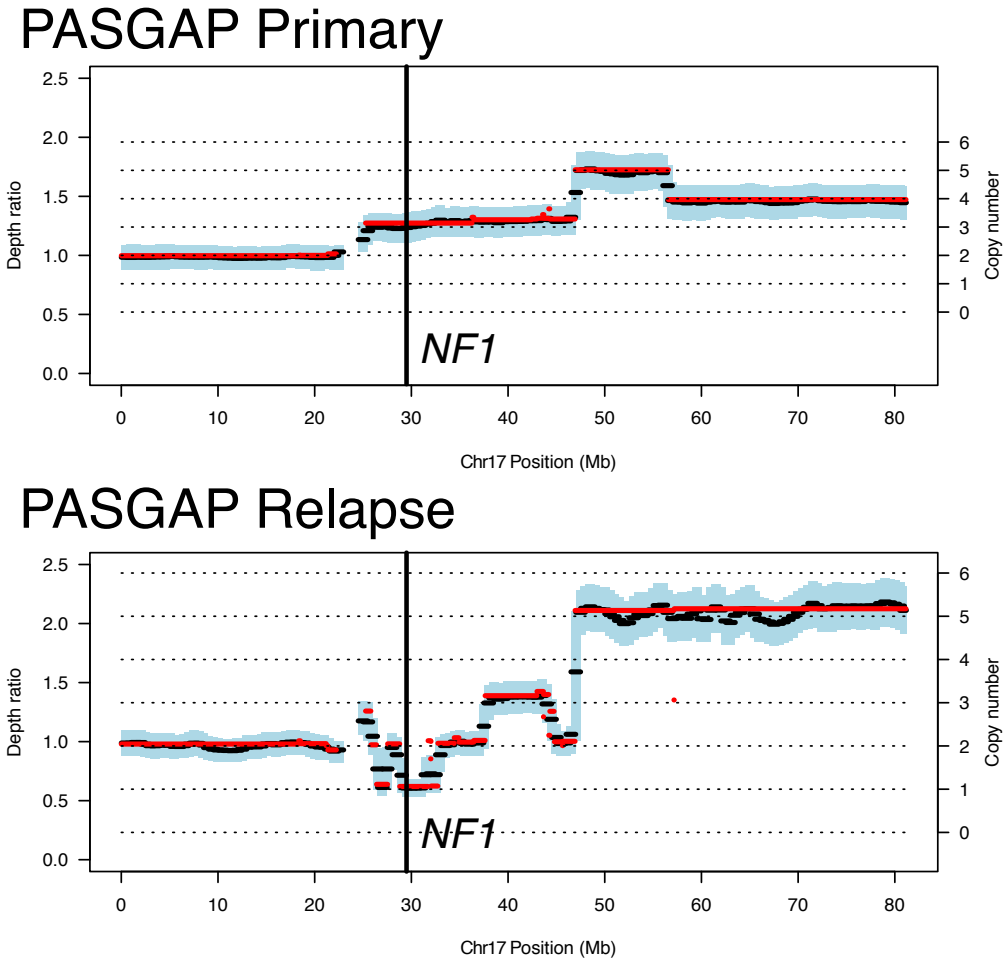

Figure 3.2: Sequenza results for the primary and corresponding relapse tumors from patient "PASGAP" at chromosome 17. These data are similar to Figure 3.1, but focus on chromosome 17 in order to better illustrate the complex rearrangements in this region. Importantly, the sub-region containing the *NF1* tumor suppressor (vertical black line) underwent a relapse-specific hemizygous deletion leading to loss of heterozygosity. We can infer that the *NF1* deletion is clonal or nearly clonal because the deleted region falls onto a dotted line (see Appendix A for rationale).

Finally, by integrating DNA WGS and RNA-seq data across primary and relapse tumors of patient PASGAP, we were able to further validate the functional significance

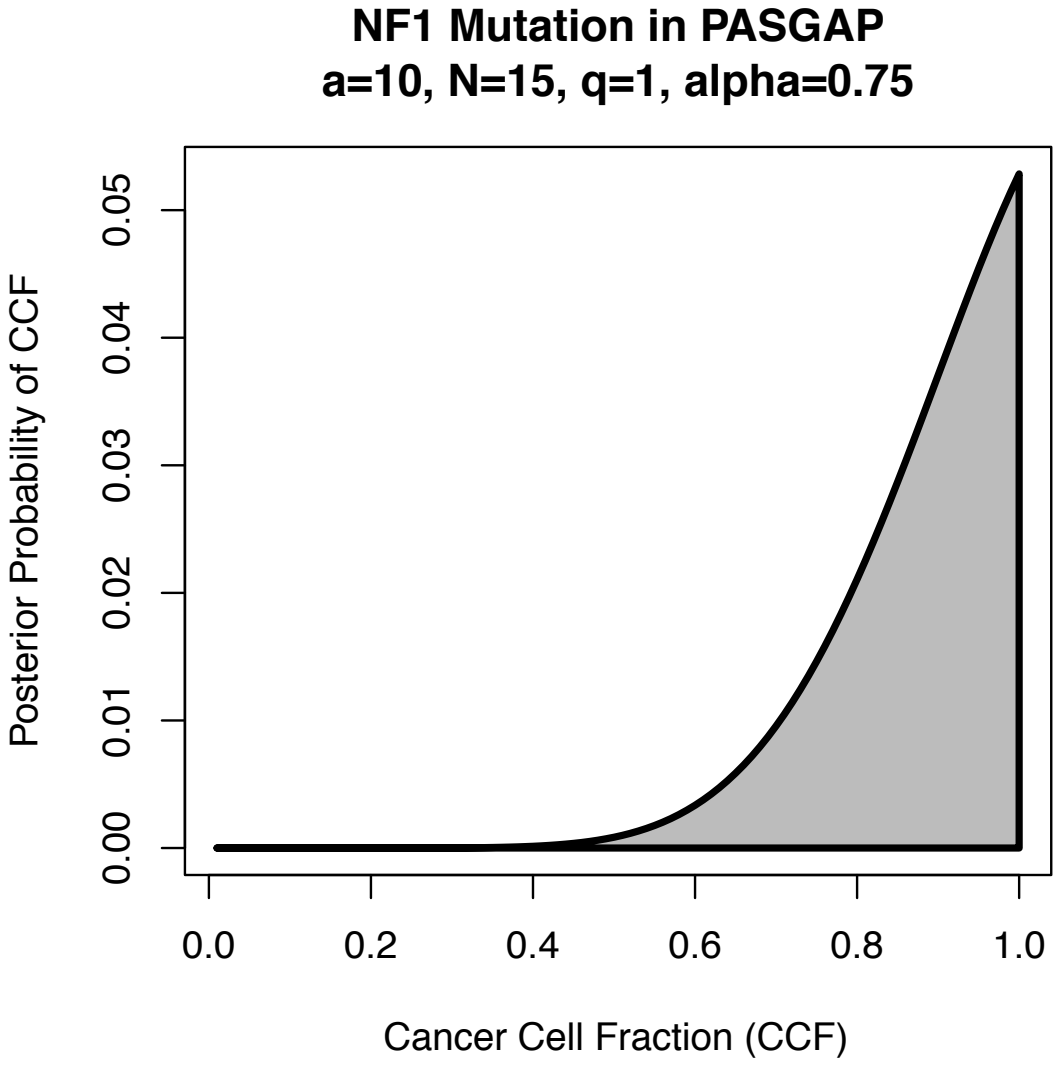

Figure 3.3: Bayesian model of cancer cell fraction of *NF1* splice site mutation in patient "PASGAP". Shown is the posterior distribution of CCF for the relapse-specific splice site mutation in PASGAP, after applying the Bayesian methodology of Carter *et al.* and Landau *et al.*[107, 108]. The posterior mode of 100% CCF suggests that this mutation is quite likely clonal or nearly clonal. A 95% credibility interval in this case is (65.5, 100%). See Appendix A for methodological details.

of these relapse-specific lesions in the *NF1* gene. In particular, we observed that the splice site mutation in PASGAP resulted in 100% expression of the mutant allele along with increased intron retention only in the relapse tumor, likely causing premature

truncation of the NF1 protein (see Figure 3.4). Thus, mutation of an *NF1* splice donor site and hemizygous deletion of the wild-type allele likely conspire to yield clonal bi-allelic inactivation of the NF1 tumor suppressor. This example illustrates one way in which RAS-MAPK pathway activation can be achieved in relapsed neuroblastomas.

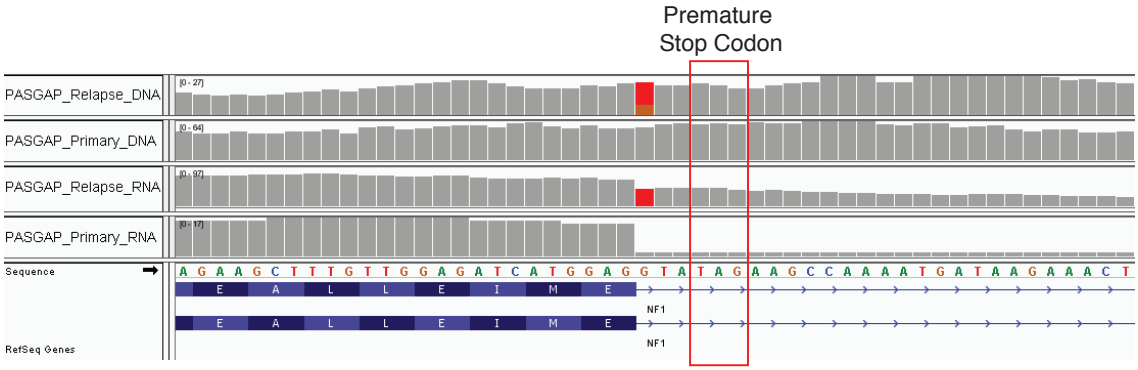

Figure 3.4: DNA and RNA sequencing profiles at the splice donor site that is mutated in the relapse tumor of patient PASGAP. Only wild-type alleles are observed in the PASGAP primary tumor, thereby maintaining in an intact splice donor site and normal splicing of *NF1*. However, in the PASGAP relapse tumor, mutation of a splice donor site results in intron retention, evidenced by increased intronic pileup whose specificity is demonstrated by 100% RNA expression of the mutant allele. The presence of an immediate in-frame stop codon within this intron likely results in a premature truncation of the NF1 tumor suppressor and corresponding loss of function.

The following section is largely copied with minor edits from our recent publication in *Nature Genetics*, whose full citation is provided here: [106]. Although this study represents the collaborative efforts of many people, my principal contributions included performing all aspects of whole genome sequencing data analysis of the American TARGET samples up until the time that we ultimately merged our efforts with our Dutch and French colleagues. During our collaboration, I continued to provide input and direct contributions to the analysis and presentation of our combined next generation sequencing data, led all aspects of the clonality analysis described in detail here, as well as contributing to many of the figures and in writing and editing the

final manuscript.

3.3 Eleveld, Oldridge, Bernard *et al.* 2015. Relapsed neuroblastomas show frequent RAS-MAPK pathway mutations.

We sequenced the genomes of 23 triplet samples constituting primary and relapse neuroblastomas and lymphocytes. Tumors were of all stages and had variable outcome, with the only eligibility criterion being the availability of high-quality DNA from the triplet samples. There was a roughly equal distribution of cases among low-, intermediate-, and high-risk groups[7]. The median time from diagnosis to relapse was 11.3 months (range of 1-90 months). Twenty-one of the 23 subjects in this study received chemotherapy before relapse, and 8 also received radiation therapy, according to internationally accepted treatment protocols (Table 3.1). Thus, all patients underwent similar chemotherapy regimens, with high-risk patients additionally receiving radiation therapy and high-dose chemotherapy with stem cell rescue. No patient on this study received targeted inhibitors to any oncogenic pathway between the time of diagnosis and relapse. The majority of low-risk cases received chemotherapy because of the site and/or size of the primary tumor.

We analyzed the sequence data for somatic mutations resulting in amino acid changes or located in splice-site regions within 3 bp of an exon, as well as for focal structural aberrations of regions containing five genes or fewer. There was a median of 14 more mutations in the relapsed samples than in the samples at diagnosis (Figures 3.5 and 3.6). On average, 28% of the mutations detected in the primary tumor were also detected at relapse, showing that the primary and relapsed tumors were of common

Table 3.1: Clinical characteristics of the relapse patient cohort.

					Time in Months				Treatment, Dx → Rel				
Patient ID	Risk Group	Stage	MYCN	Sex	Dx	Rel	Last	Status	Rad	Chem	Surg	Dx Location	Rel Location
FR_NB1178	High	4	Non-amplified	M	30	21	24	Dead	Y	Y	Y	Retroperitoneum	Liver
FR_NB1269	High	4	Amplified	M	14	9	11	Dead	N	Y	Y	Retroperitoneum	Retroperitoneum
FR_NB1382	High	4	Amplified	M	4	50	64	Dead	N	Y	Y	Abdomen	Abdomen
NL_N774	High	4	Non-amplified	F	83	6	10	Dead	Y	Y	Y	Adrenal gland	Abdomen
US_PATNKP	High	4	Non-amplified	M	113	20	40	Alive	Y	Y	Y	Retroperitoneum	Pelvis
US_PASGAP	High	4	Non-amplified	M	44	42	51	Dead	Y	Y	Y	Adrenal gland	Soft tissue, skull
NL_N790	High	3	Amplified	F	42	63	110	Dead	Y	Y	Y	Adrenal gland	Liver
US_PASHFA	High	3	Amplified	F	13	7	11	Dead	N	Y	Y†	Adrenal gland	Abdomen
FR_NB804	Intermediate	4	Non-amplified	F	2	26	56	Alive	N	Y	Y	Subcutaneous nodule	Orbita
NL_N607	Intermediate	4	Non-amplified	F	3	7	84	Alive	Y	Y	N	Liver	Orbita
US_PARHAM	Intermediate	4	Non-amplified	F	11	1	81	Dead	N	Y	N	Pelvis	Pelvis
US_PATYIL	Intermediate	4	Non-amplified	F	11	8	16	Dead	N	Y	Y	Abdomen	Pararenal
NL_N571	Intermediate	3	Non-amplified	M	49	12	16	Dead	N	Y	Y	Adrenal gland	Abdomen
US_PASNPG	Intermediate	3	Non-amplified	F	10	10	63	Alive	N	Y	Y	Retroperitoneum	Paraspinal
US_PARBAJ	Intermediate	3	Non-amplified	M	1	10	88	Alive	N	Y	Y	Retroperitoneum	Abdomen
US_PAUDDK	Intermediate	3	Non-amplified	M	12	11	38	Alive	N	Y	Y†	Pelvis	Pelvis
FR_NB1224	Low	2	Non-amplified	M	15	8	18	Alive	N	Y	N	Mediastinum	Mediastinum
FR_NB0175	Low	2	Non-amplified	M	98	90	103	Dead	N	N	Y	Retroperitoneum	Retroperitoneum
FR_NB308	Low	2	Non-amplified	F	2	21	91	Alive	N	Y	N	Abdomen	Abdomen
NL_N041	Low	2	Non-amplified	F	109	72	92	Dead	Y	Y	Y	Abdomen	Abdomen
US_PAPVEB	Low	2	Non-amplified	M	57	9	40	Dead	N	N	N	Adrenal gland	Bone marrow
NL_N789	Low	1	Non-amplified	F	124	78	168	Alive	Y	Y	Y	Adrenal gland	Lymph node
FR_NB399	Low	4s	Non-amplified	M	0	7	134	Dead	N	Y	N	Subcutaneous nodule	Liver

Dx = diagnosis; Rel = relapse; Rad = radiation; Chem = chemotherapy; Surg = surgery
†Biopsy only

descent. To gain more insight into the clonal architecture, we estimated the cancer cell fraction (CCF) of all somatic mutations using a customized reimplementation of a previously described Bayesian approach[107], which infers CCF from mutant allele fractions determined by sequencing and accounts for normal contamination and locus-specific copy number. This analysis yielded a median CCF of 61% for mutations detected in the primary tumor but lost in the relapse tumor, in comparison to a median CCF of 90% for primary tumor mutations shared with the relapse tumor (Figure 3.7a), a pattern that is consistent with subclonal outgrowth of the relapsed tumor. Furthermore, we estimated a median CCF of 83% for all relapse tumor mutations in comparison to a median CCF of 75% for all primary tumor mutations, indicating clonal enrichment of a subset of mutations at relapse (Figure 3.7b). Comparison of genes affected by smaller structural events and chromosomal copy number alterations in primary and paired relapse tumors showed similar results, with a subset of aberrations shared but many detected only in the primary or relapse tumor (Figure 3.8).

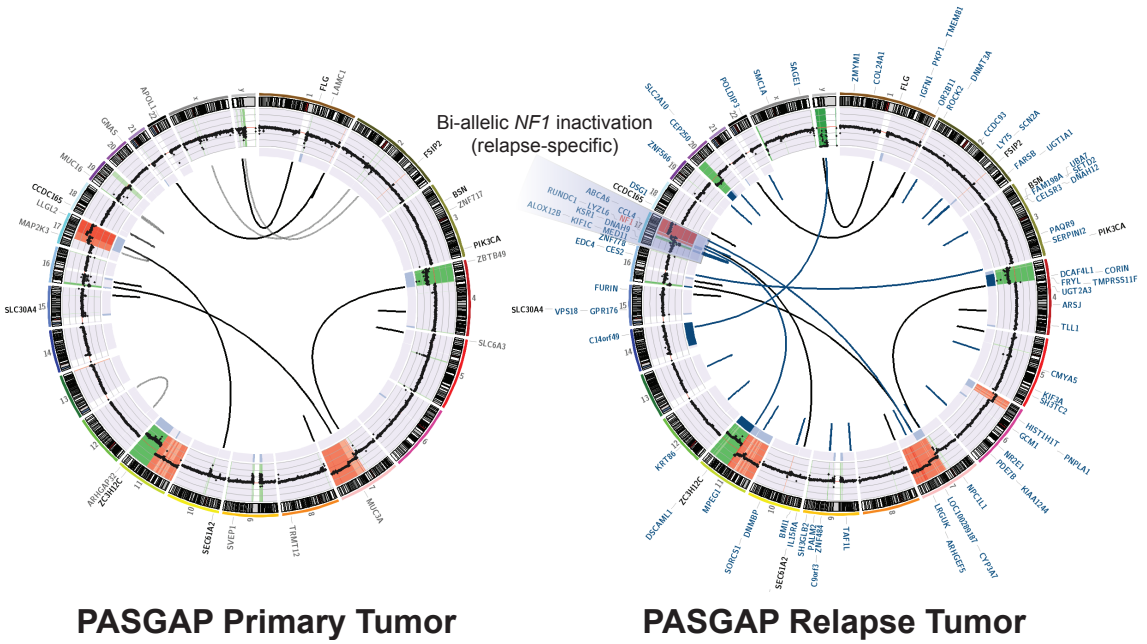

Figure 3.5: Circos plots showing structural alterations and somatic mutations in the primary and corresponding relapse tumors from PASGAP. The inner ring represents copy number alterations (red, gain; green, loss) identified on the basis of coverage of the tumor and lymphocyte genomes. The lines traversing the ring indicate inter- and intrachromosomal rearrangements identified by discordant mate pairs from paired-end reads. Aberrations are colored according to their presence (gray, only detected in the primary tumor; black, detected in the primary and relapse tumors; blue, only detected in the relapse tumor; red, events predicted to activate RAS-MAPK signaling). PASGAP exhibits bi-allelic inactivation of *NF1* that is only observed at relapse.

Enrichment of mutations that activate RAS-MAPK signaling

Unbiased pathway analysis[110] using whole-genome sequencing data from the relapse samples on a per-patient basis identified a strong enrichment ($P = 6.1 \times 10^{-7}$) for mutations in genes associated with RAS-mitogen-activated protein kinase (MAPK) signaling. We next filtered the identified mutated genes against the Cancer Gene Census[111] and subsequently focused on hotspot regions by selecting mutations annotated in the Catalogue of Somatic Mutations in Cancer (COSMIC) to identify events that are well annotated to activate this pathway. Fifteen of 23 relapse samples

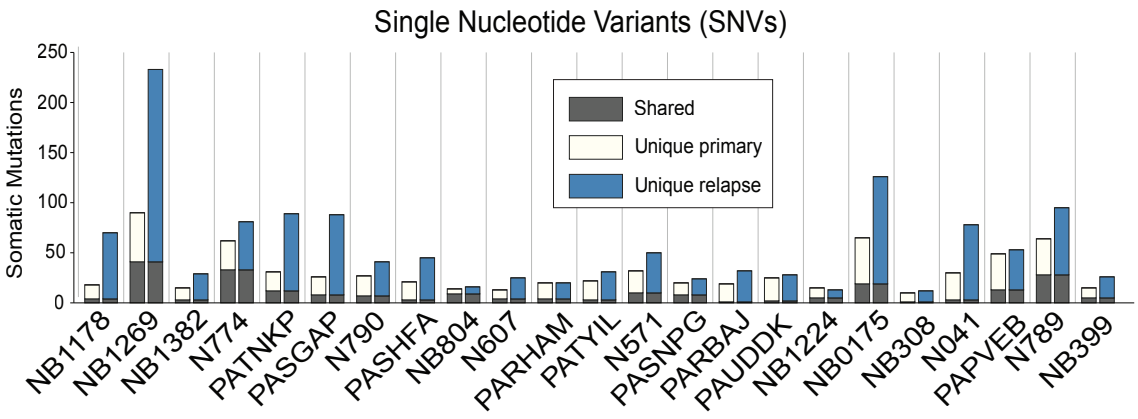

Figure 3.6: Count of nonsynonymous mutations identified by whole-genome sequencing in 23 primary tumors and their corresponding relapse tumors. Mutations identified in both primary and relapse tumors are shown in gray, whereas mutations unique to the primary or relapse tumor are shown in beige and blue, respectively. On average, more mutations were observed at relapse, indicating clonal evolution.

contained somatic mutations that met these criteria. In addition, three relapse samples showed structural alterations involving these RAS-MAPK pathway genes; thus, we detected aberrations in this pathway in 18 of 23 relapse samples (78%), and all alterations were consistent with pathway activation (Table 3.2). Eleven mutations activating RAS-MAPK signaling that were present in primary tumors were all preserved in the corresponding relapse tumors. Seven mutations were not detectable in the primary tumor at the sequencing depth achieved, and we therefore employed ultra-deep sequencing to determine whether these mutations were present in fractions under the whole-genome sequencing detection limit. The *ALK* mutation in N607 was found in the primary tumor, whereas the other mutations were undetectable. We assayed structural aberrations using PCR-based methods. Only the ALK rearrangement in N790 was shown to be present at low frequency in the primary tumor.

ALK aberrations occurred in ten relapse tumors and were also detected by whole-genome sequencing in seven of the corresponding primary tumors. All detected single-

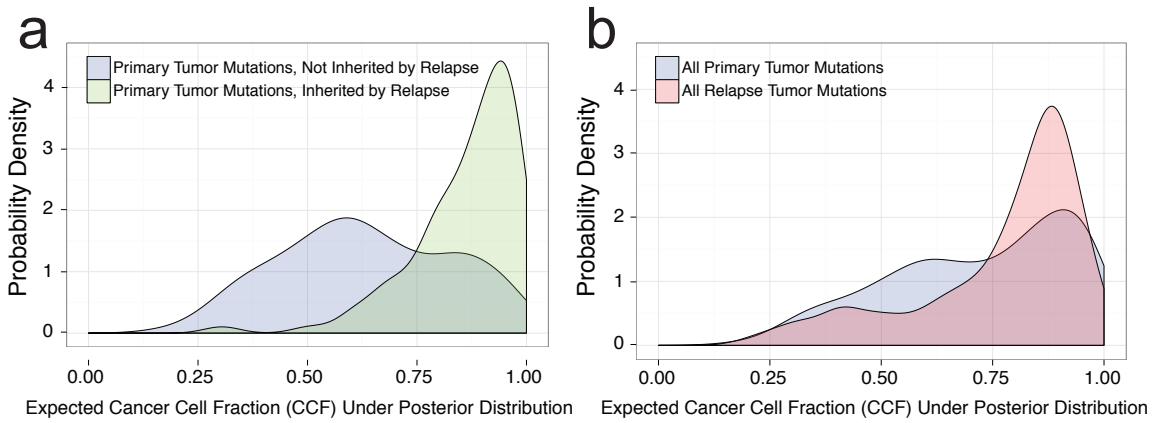

Figure 3.7: Relapse tumors undergo clonal evolution over the course of treatment. The posterior probability distribution of cancer cell fraction (CCF) was estimated for all somatic coding SNVs by the method of Carter *et al.*[107] assuming a somatic multiplicity of one. The expected value of CCF for each SNV, E[CCF], was then computed under its respective posterior distribution, and the empirical density of E[CCF] was computed across all somatic coding SNVs and patients and visualized in the above plots. (**a**) Comparison of primary tumor mutations that are inherited by their corresponding relapse tumor versus those that are lost. (**b**) Comparison of all primary tumor mutations versus all relapse mutations. The enrichment of high-CCF mutations at relapse and the loss of low-CCF mutations detected in the primary tumor are consistent with subclone outgrowth and clonal enrichment of a subset of mutations at relapse.

nucleotide variants (SNVs) have been proven to constitutively activate this receptor tyrosine kinase known to activate RAS-MAPK signaling[112]. Furthermore, one relapse sample showed a *de novo* amplification giving rise to a *PPM1G-ALK* fusion gene, which activated the RAS-MAPK pathway when expressed in neuroblastoma cell lines.

Two tumors showed relapse-specific inactivation of the *NF1* tumor-suppressor gene, through homozygous deletion in one case and heterozygous deletion combined with a splice-site mutation in the other. *NF1* inactivation has been reported in neuroblastoma and confers activation of RAS-MAPK signaling and resistance to retinoic acid[113]. One pair of primary and relapse tumors showed a heterozygous mutation in *PTPN11*. Mutations in *PTPN11* activate RAS-MAPK signaling, and the

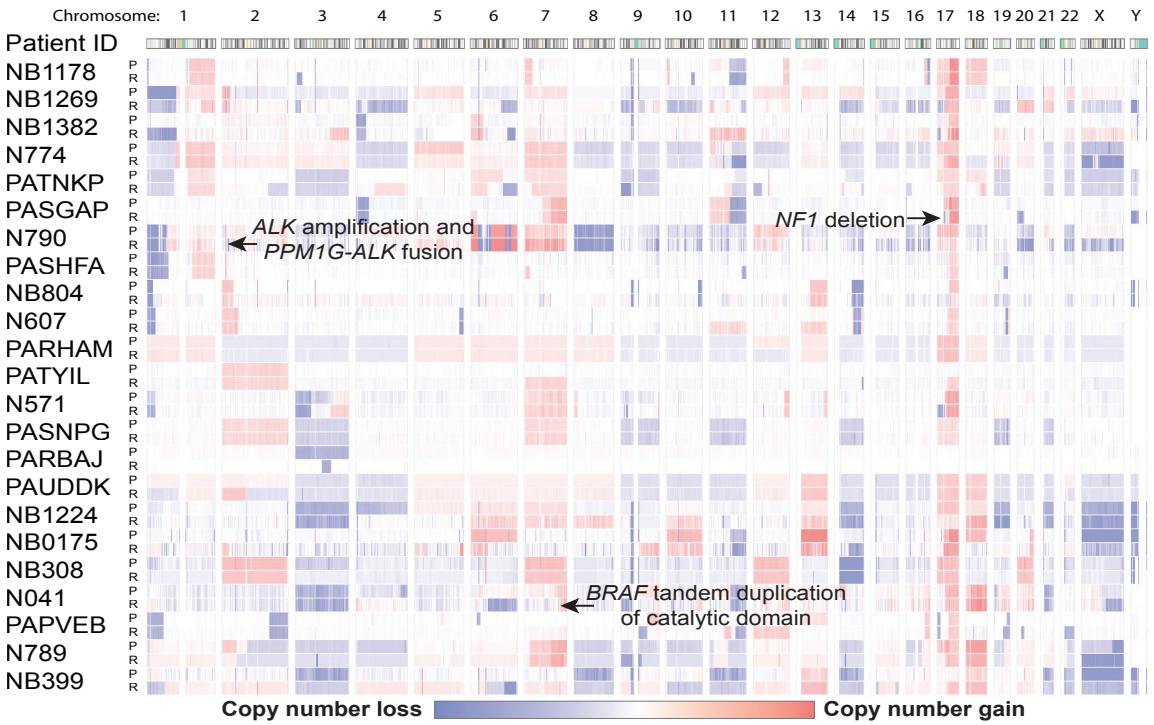

Figure 3.8: Relative coverage plots displaying the structural alterations in primary and relapse tumors. Coverage was calculated for 1-Mb bins along the genome and normalized to the coverage in patient lymphocyte DNA. Color intensity reflects the magnitude of the gain or loss (blue, loss; red, gain) in the associated chromosomal location. Focal RAS-MAPK-pathway-activating structural alterations are labeled with arrows. P = primary; R = relapse

identified mutation encoding a p.Ala72Thr substitution has been reported in leukemia and neuroblastoma[114].

One relapse tumor showed a tandem duplication in the *BRAF* gene that was not detected in the primary tumor. This rearrangement leads to expression of a *BRAF* transcript that encodes an elongated protein with two kinase domains. Expression of this *BRAF* gene with a tandem duplication in a neuroblastoma cell line induced activation of the RAS-MAPK pathway. Taken together, the somatic mutations detected in this case series shown to activate the RAS-MAPK pathway were mutually exclusive, with no case having two somatic events known to hyperactivate this growth-promoting

Table 3.2: RAS-MAPK pathway mutations in relapsed neuroblastomas.

		Genomic aberrations in RAS-MAPK pathway in relapse tumors			
Patient ID	Gene	Genomic event	Event type	COSMIC ID	In primary†
FR_NB1178					
FR_NB1269	*ALK*	Somatic mutation (L1196M)	Activating	99137	Yes
FR_NB1382	*ALK*	Somatic mutation (Y1278S)	Activating	28058	No
NL_N774	*PTPN11*	Somatic mutation (A72T)	Activating	13014	Yes
US_PATNKP	*FGFR1*	Somatic mutation (N546K)	Activating	19176	Yes
US_PASGAP	*NF1*	Somatic Mutation (Splice Donor) + Hemizygous Deletion	Inactivating		No
NL_N790	*ALK*	Amplification and fusion	Activating		No
US_PASHFA					
FR_NB804					
NL_N607	*ALK*	Somatic mutation (F1174L)	Activating	28055	No
US_PARHAM	*ALK*	Somatic mutation (R1275Q)	Activating	28056	Yes
US_PATYIL	*NRAS*	Somatic mutation (Q61K)	Activating	580	Yes
NL_N571	*NF1*	Homozygous deletion	Inactivating		No
US_PASNPG	*ALK*	Somatic mutation (F1174I)	Activating	28491	Yes
US_PARBAJ	*HRAS*	Somatic mutation (Q61K)	Activating	496	No
US_PAUDDK					
FR_NB1224	*ALK*	Somatic mutation (R1275Q)	Activating	28056	Yes
FR_NB0175	*ALK*	Somatic mutation (Y1278S)	Activating	28058	Yes
FR_NB308	*ALK*	Somatic mutation (F1174L)	Activating	28061	Yes
NL_N041	*BRAF*	Tandem duplication catalytic domain	Activating		No
US_PAPVEB	*KRAS*	Somatic mutation (G12D)	Activating	521	Yes
NL_N789					
FR_NB399	*ALK*	Somatic mutation (R1275Q)	Activating	28056	Yes

†Detected in primary tumor by whole genome sequencing analysis

pathway.

We hypothesized that mutations activating RAS-MAPK signaling exhibit relapse-specific enrichment due to treatment. We therefore performed a clonality analysis, comparing CCF estimates for RAS-MAPK mutations in paired primary (CCF_p) and relapse (CCF_r) tumors (Figure 3.9). RAS-MAPK mutations were almost universally present within major subclonal populations at relapse, as indicated by $\text{CCF}_r > 0.5$ with probability $> 90\%$ under the posterior distribution for 14 of 15 relapse tumors. In 7 of 15 tumor pairs, there was strong evidence of relapse-specific enrichment of RAS-MAPK mutations—including mutations in *ALK* (4 pairs), *HRAS* (1 pair), *KRAS* (1 pair) and *NF1* (1 pair)—based on a criterion of $\text{CCF}_r > \text{CCF}_p$ with probability $> 90\%$ for each pair. By contrast, the probability that $\text{CCF}_r > \text{CCF}_p$ fell within 20-80% bounds for the remaining eight pairs, indicating that RAS-MAPK mutations that were already present in the primary tumor were retained at relapse. Collectively,

these results support RAS-MAPK mutations as somatic drivers that undergo positive selection over the course of neuroblastoma treatment.

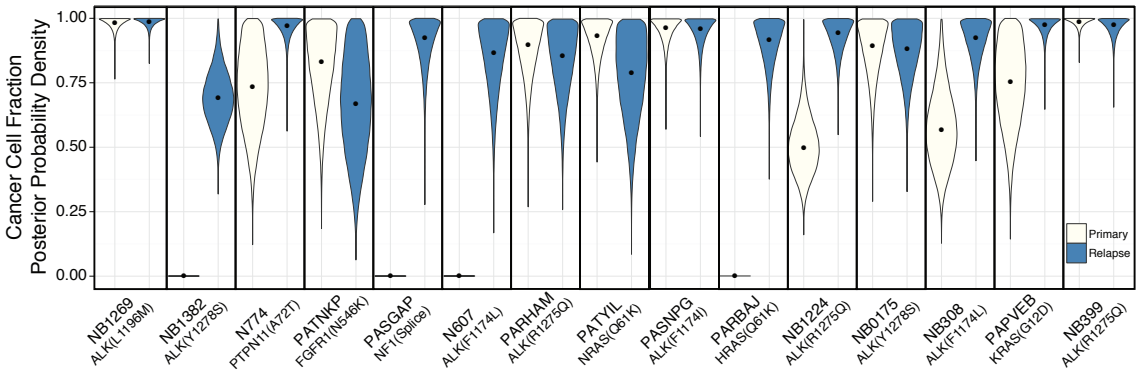

Figure 3.9: RAS-MAPK pathway mutations reside within major relapsed neuroblastoma subclones. Each of the 15 panels represents a primary-relapse pair with a corresponding RAS-MAPK pathway mutation. Posterior distributions of CCF were computed by the method of Carter and colleagues[107] and are represented by violin plots, with black dots positioned at the distribution medians. Four primary-relapse pairs (NB1382, PASGAP, N607 and PARBAJ) possess relapse-specific RAS-MAPK pathway mutations that are undetectable in the corresponding primary tumors by whole-genome sequencing. Three additional pairs (NB1224, NB308 and PAPVEB) also show evidence of relapse-specific enrichment of RAS-MAPK mutations based on the criterion that the probability of $\mathrm{CCF}_r > \mathrm{CCF}_p$ was $> 90\%$ for each pair. CCF estimates for *ALK* mutations in NB308 reflect p.Phe1174Leu (c.352C→A) and p.Phe1174Leu (c.352C→G) substitutions in the primary and relapse samples, respectively, as previously described[104].

Chromosomal aberrations

Three relapse samples showed homozygous deletions in the *CDKN2A* locus, encoding the tumor-suppressor proteins p14ARF and p16, whereas both alleles of *CDKN2A* were present in the corresponding primary tumors. *CDKN2A* deletions were previously reported as frequent events in neuroblastoma relapse[115]. We detected other relapse-specific segmental chromosome defects, including loss of 6q (five cases) and loss of 17p (three cases). Furthermore, we detected relapse-specific aberrations that are frequently detected in primary neuroblastoma and are associated with poor prognosis, including loss of chromosomes 1p (one case) and 11q (three cases) (Figure 3.8)[18, 20].

RAS-MAPK pathway mutations and sensitivity to MEK inhibition

To determine whether neuroblastoma cell lines contain RAS-MAPK mutations, we analyzed whole-genome sequencing data from a series of human-derived neuroblastoma cell lines for mutations in *ALK*, *NRAS*, *HRAS*, *KRAS*, *BRAF*, *PTPN11* and *NF1*. Eleven of the 18 cell lines showed such mutations.

We tested our cell line panel for sensitivity to the MEK inhibitors trametinib, cobimetinib and binimetinib, to determine the relationship between mutation status and drug sensitivity. The data showed clustering of the cell lines into four groups with increasing sensitivity to MEK inhibition: (i) lines without RAS-MAPK mutations; (ii) lines with *ALK* mutations; (iii) lines with *NF1* mutations; and (iv) lines with *RAS* gene or *BRAF* mutations. In the cell lines with mutated *RAS* genes or *BRAF*, MEK inhibitor treatment caused almost complete cell cycle arrest at low nanomolar concentrations, whereas in the *NF1*- and *ALK*-mutated lines the effect on cell cycle inhibition was less robust (data not shown). When the sensitivity of the cell lines was expressed as the concentration at which cell growth was inhibited by 50% (GI_{50}), there were significant differences between cell lines with and without RAS-MAPK mutations (Figure 3.10). GI_{50} values for the three different compounds were highly correlated in the cell line panel, suggesting an on-target effect (r^2 = 0.49-0.79; $P < 0.01$). A relationship between mutation status and sensitivity to MEK inhibition was also observed in an independent published data set[116].

To validate that *ALK* and *RAS* gene mutations directly activate the RAS-MAPK pathway in neuroblastoma cells, we induced the expression of an *ALK* Phe1174Leu and an *NRAS* Gln61Lys mutant in two cell lines that did not harbor RAS-MAPK mutations. Expression of either mutated protein caused activation of the pathway. We have shown previously that knockdown of *NF1* causes hyperactivated RAS-MAPK signaling in neuroblastoma cell lines[113].

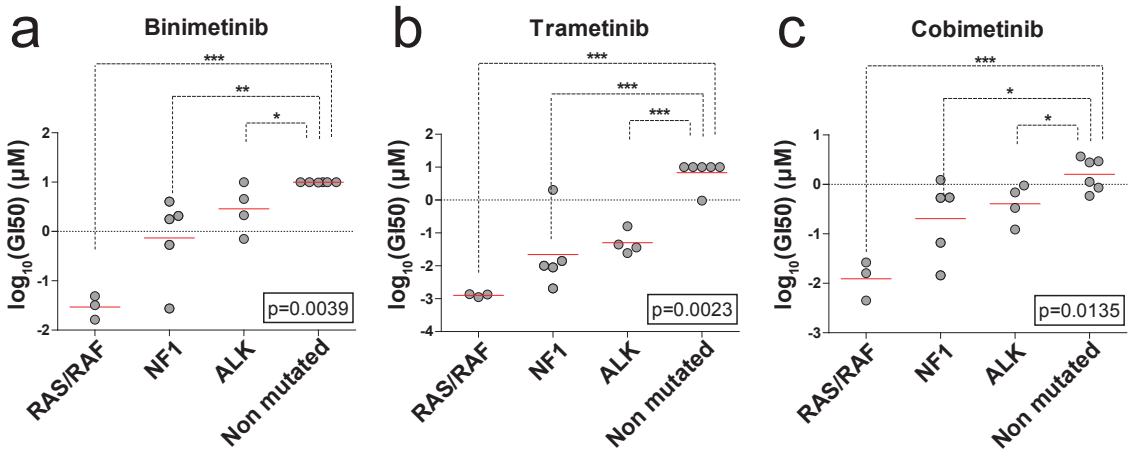

Figure 3.10: Sensitivity of neuroblastoma *in vitro* cell line models to MEK inhibition therapy. (**a-c**) GI_{50} values for a panel of neuroblastoma cell lines treated with binimetinib (**a**), trametinib (**b**) and cobimetinib (**c**). Cell lines are grouped according to mutation status. Red bars represent the mean GI_{50} value for each group. *P*-values were derived from Kruskal-Wallis tests. Student's *t*-tests were performed to determine differences between non-mutated and mutated groups: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

We next treated various human neuroblastoma-derived cell line xenograft models, representing the four groups described above, with the MEK inhibitor binimetinib. SK-N-AS xenografts, which harbor an *NRAS* mutation encoding p.Gln61Lys, showed inhibition of tumor growth and increased survival when treated with binimetinib, in a dose-dependent fashion (Figure 3.11). NBL-S xenografts have an inactivating mutation in one allele of *NF1* and an almost complete absence of *NF1* protein expression; these xenografts also showed inhibition of growth with treatment. Conversely, treatment of Kelly and IMR-5 xenografts had no effect on tumor growth. IMR-5 cells do not have RAS-MAPK pathway mutations detectable by whole-exome sequencing (data not shown), whereas Kelly cells harbor an *ALK* mutation encoding a p.Phe1174Leu substitution.

We then determined whether inhibition of cell growth corresponds with inhibition of the RAS-MAPK pathway in the cell lines that were used for the mouse xenograft

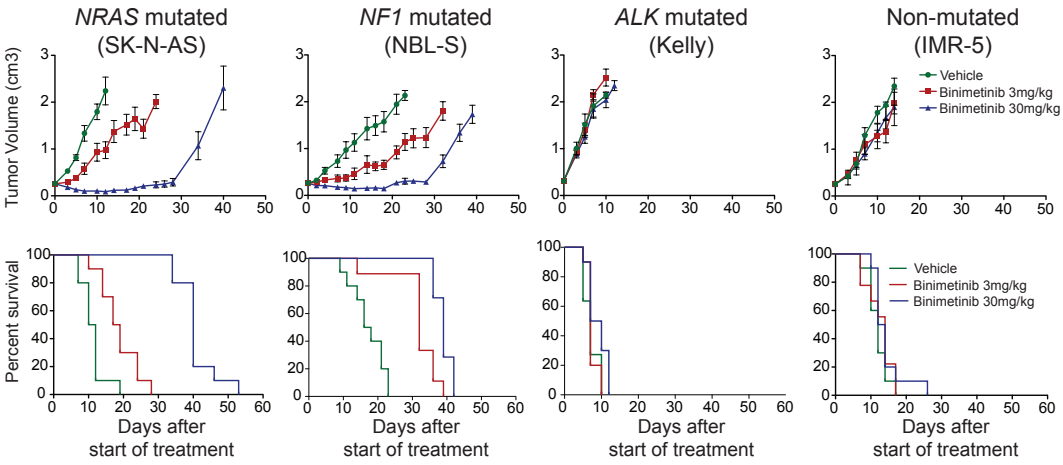

Figure 3.11: Sensitivity of neuroblastoma cell line-derived xenograft models to MEK inhibition therapy. Human neuroblastoma-derived SK-N-AS, NBL-S, Kelly and IMR-5 xenografts were treated with binimetinib (3 mg/kg or 30 mg/kg) or vehicle by oral gavage twice daily. Each cohort consisted of ten mice, and tumor volumes (cm^3) and percent survival are shown. Error bars denote ± s.e.m.; significance is denoted: $^*P < 0.05$.

experiments. We treated the cell lines with increasing concentrations of binimetinib *in vitro* for 24 h and screened for ERK phosphorylation. The three lines with RAS-MAPK mutations showed phosphorylation of ERK under untreated conditions, reaffirming that these mutations indeed lead to activation of the RAS-MAPK pathway. Upon exposure to binimetinib, SK-N-AS and NBL-S cells showed a dose-dependent decrease in the levels of phosphorylated ERK, whereas Kelly cells showed no significant change. These data suggest that the minimal effect of MEK inhibition *in vitro* and the absence of an effect *in vivo* for the Kelly cell line may be due, at least in part, to continued ERK phosphorylation. To confirm that the response we observed *in vivo* was also due to target inhibition, we analyzed ERK phosphorylation in the xenografts that responded to treatment and demonstrated a dose-dependent decrease in the levels of phosphorylated ERK.

3.4 Summary and future directions

3.4.1 Clinical significance and the need for relapse biopsies

In this study, we characterized the genomes of 23 relapsed neuroblastomas and compared each to the genome of the corresponding primary tumor. We show that the relapsed tumors generally contain more mutations and structural aberrations and that clonal selection takes place between the primary tumor and the relapse tumor. We found that 18 of 23 relapse tumors harbored mutations predicted to hyperactivate the RAS-MAPK signaling pathway and that cell lines containing similar mutations show sensitivity to inhibition of MEK, a downstream node in the canonical growth-promoting pathway. These results provide a strong rationale for recommending biopsy and genomic characterization of relapsed neuroblastoma tumors and for prioritizing the clinical testing of MEK inhibition strategies in the treatment of relapsed neuroblastoma.

Because of the sensitivity and specificity of modern imaging modalities, the diagnosis of neuroblastoma relapse rarely requires a tumor biopsy. In addition, until recently, there has been no realistic potential for therapeutic benefit based on biopsy results, and therefore very few relapse neuroblastoma samples are available for study. Here we collected high-quality material from 23 primary-relapse tumor pairs across the spectrum of neuroblastoma phenotypes, including those assigned to high-, intermediate- and low-risk groups[7]. The only inclusion criterion for this study was the availability of matched samples, but some cases may have been biopsied owing to an unusual clinical course. Indeed, there were a fairly high number of intermediate- and low-risk tumors that normally do not show frequent relapse[7]. However, the frequency of RAS-MAPK mutations did not differ among the groups, so it is unlikely that the over-representation analysis is influenced by this bias.

The high frequency of RAS-MAPK pathway mutations at diagnosis in this cohort

was unexpected, as such high frequencies were not reported in whole-genome sequencing series of primary tumors[44–47]. It is possible that the presence of these lesions in a diagnostic sample is a biomarker of a more aggressive clinical course and a higher likelihood of relapse. These findings need to be validated in a prospective patient cohort.

We detected several recurrent structural aberrations at the time of relapse. Partial loss of chromosome 6q was observed in five relapse samples, and homozygous deletion of *CDKN2A* was found in three relapse samples. Both events are infrequent in primary neuroblastoma and present interesting targets for further research. Furthermore, neuroblastoma-associated aberrations such as loss of 1p and 11q were observed in the relapse tumor and not in the primary tumor, indicating that these events might not be tumor initiating but rather are crucial steps in neuroblastoma tumor evolution[117].

Events affecting RAS-MAPK signaling were detected in 18 of 23 relapse samples. In four cases, we identified structural variants, highlighting the benefit of whole-genome sequencing for detection of the full spectrum of genetic alterations. In 7 of these 18 cases, the mutations were observed only in the relapse tumor, which indicates that analysis of primary tumor samples is not sufficient to guide the choice of treatment for neuroblastoma relapses. These findings are in line with the *de novo* occurrence of *ALK* mutations reported previously[104].

The observation that several RAS-MAPK mutations were present in the relapse tumor but not in the corresponding primary tumor favors a model where subclones with secondary driver mutations expand over time, possibly under the selective pressure of chemotherapy, as was recently described for chronic lymphocytic leukemia[108]. Whether these mutations occurred between diagnosis and relapse, were present at levels below detection limits or were undetectable owing to spatial heterogeneity in the primary tumor remains to be determined.

It has been firmly established that mutations in the RAS-MAPK pathway can occur as resistance mechanisms against treatment with targeted kinase inhibitors[118]; however, no targeted inhibitors were used in the treatment of our patient cohort. Mutations in the RAS-MAPK pathway may also be associated with resistance to conventional cytotoxic therapies in neuroblastoma, but more research is needed to establish the molecular basis of this phenotype.

The *ALK* gene was mutated in ten relapse samples, and it is known that the most frequent *ALK* mutations in neuroblastoma activate the RAS-MAPK signaling pathway[39]. The results of our xenograft therapeutic studies indicate that single-agent treatment with a MEK inhibitor might not be effective in *ALK*-mutated tumors. However, the finding that ALK-mutated cell lines consistently showed some sensitivity to MEK inhibition *in vitro* suggests that activated RAS-MAPK signaling does have a role in *ALK*-mutated neuroblastoma and warrants further investigation on the use of MEK inhibitors in combination therapies. ALK inhibitors have proven to be effective in the treatment of *ALK*-mutated tumors[43], but some mutations are associated with resistance to currently available ALK inhibitors[112]. Therefore, combined MEK and ALK inhibition might improve response in tumors containing such mutations. Combination of these inhibitors with ones targeted against other pathways that are activated in *ALK*-mutated neuroblastoma, such as phosphoinositide 3-kinase (PI3K) and mTOR[119, 120] signaling, might also improve therapeutic efficacy.

We also detect mutations in *NF1*, *BRAF*, *PTPN11*, *FGFR1* and the three *RAS* genes, and all lesions in the RAS-MAPK pathway were mutually exclusive. Cell lines with RAS-MAPK mutations show moderate to high sensitivity to MEK inhibitors, and treatment with the inhibitor binimetinib of SK-N-AS xenografts, which contain an *NRAS* mutation, as well as NBL-S cells, which have loss of *NF1*, results in significant therapeutic efficacy. These findings suggest that MEK inhibition might be of clinical

benefit in the treatment of neuroblastoma relapses containing RAS-MAPK mutations.

3.4.2 NEPENTHE: a phase 1b/2 clinical trial for relapsed neuroblastoma

Our observation of RAS-MAPK pathway-activating mutations in relapsed neuroblastomas has provided part of the rationale for NEPENTHE, a soon to open pediatric cancer phase 1b/2 clinical trial that will match genomic aberrations in relapsed tumors to combinations of targeted therapeutics that were identified through synergy screens. Subjects will be enrolled into the following treatment groups depending on their mutation profiles:

Group 1: Subjects with activating mutations in *ALK* will receive Ceritinib (ALK-inhibitor) + LEE011 (CDK4/6-inhibitor). Ceritinib is a second generation ALK inhibitor that is more potent than crizotinib, has elicited clinical responses in crizotinib-resistant non-small cell lung cancer[121], and has shown efficacy in pre-clinical neuroblastoma models for crizotinib-resistant ALK mutations including F1174L. Dual CDK4/6 inhibition has been shown to induce cell-cycle arrest and senescence in neuroblastoma cell lines and cell line derived xenografts[122], and unpublished pre-clinical data indicate that combined ALK- and CDK4/6-inhibition is synergistic.

Group 2: Subjects with somatic *MYCN* amplification and/or RAS/MAPK pathway lesions and/or cell cycle regulatory gene lesions will receive Trametinib (MEK-inhibitor) + LEE011 (CDK4/6-inhibitor), a combination which has shown synergy in pre-clinical models.

Group 3: Subjects that do not match groups 1 or 2 and whose tumors show wild-type *TP53* will receive single-agent HDM201 (MDM2-inhibitor). While the majority of neuroblastomas have intact p53, increased activity of its principal negative regulator,

MDM2, may lead to aberrant p53 signaling in neuroblastoma and present a therapeutic opportunity[123, 124].

In summary, NEPENTHE represents the first trail of combination therapy that is targeted to specific mutational profiles that we observed in our recently published study of relapsed neuroblastoma genomes. Additionally, NEPENTHE will provide an important opportunity to profile a large number of relapse tumor genomes prospectively, addressing important sources of bias from our initial study that will yield a more comprehensive view of the genomic landscape of relapsed neuroblastoma.

3.4.3 Schramm *et al.* 2015

Concurrent with our study, our German colleagues conducted a similar investigation on an independent cohort of 16 paired primary and relapsed tumors by whole genome sequencing, published back-to-back with our manuscript in *Nature Genetics* earlier this year[125]. While 2 out of 16 tumors harbored RAS mutations—in *HRAS* and *KRAS*, respectively—the overall prevalence of RAS-MAPK mutations was considerably less than the 18 out of 23 we observed. Recurrent alterations at relapse included mutations in the putative *CHD5* neuroblastoma tumor suppressor (2 patients), chromosome 9p losses (5 patients), *DOCK8* (6 patients: 2 copy number loss + 4 mutation), inactivating mutations in *PTPN14* (2 patients) and a relapse-specific activity pattern for the PTPN14 target YAP identified by differential expression profiling[125].

Differences between clinical populations are likely to account for some of the observed differences between our Eleveld *et al.* study versus the Schramm *et al.* study Indeed, whereas 21 out of 23 patients received chemotherapy prior to relapse in our combined American/Dutch/French cohort[106], it appears that this was the case for only 6 out 16 patients in the German study[125]. Chemotherapy is very

likely to exert significant selective pressure on neuroblastoma tumors (indeed, we see evidence of this in Figures 3.7 and 3.9), thereby driving which mutations appear at relapse. This is an intriguing hypothesis that warrants further study, and data collected as part of the NEPENTHE trial may shed additional light on this. However, recent reports that YAP can functionally replace KRAS as a bypass mechanism in KRAS-addicted cancer[126, 127] opens the possibility that the RAS-MAPK and YAP pathways identified in our respective studies may in fact be converging on a common pathway.

3.4.4 The clonal landscape of neuroblastoma

The insight that sub-clonal mutations present in the primary neuroblastoma tumors can undergo significant clonal enrichment at relapse has prompted us to revisit our analysis of primary tumor sequencing data generated as part of the TARGET project, including whole exome sequencing data for 222 primary tumor/blood normal pairs and targeted gene panel sequencing data for an additional 500 pairs. In order to achieve higher variant-calling specificity, the initial analysis used an effective variant allele fraction (VAF) cutoff of approximately 10% for variant calling, which can miss low frequency variants. I have now been able to rerun an analysis of the TARGET data simply lowering the VAF threshold to 5%, which identified additional mutations in important neuroblastoma oncogenes such as *ALK*, *KRAS*, and *NRAS* (see Figure 3.12). In fact, our initial estimates of *ALK* mutation prevalence at 8-9% may need to be revised upward to at least 10-11%, after accounting for low frequency variants.

We are now designing an ultra-deep targeted sequencing panel in order to better define the limit of detection for low frequency mutations and to more accurately estimate the prevalence of clonal and subclonal somatic driver mutations at diagnosis.

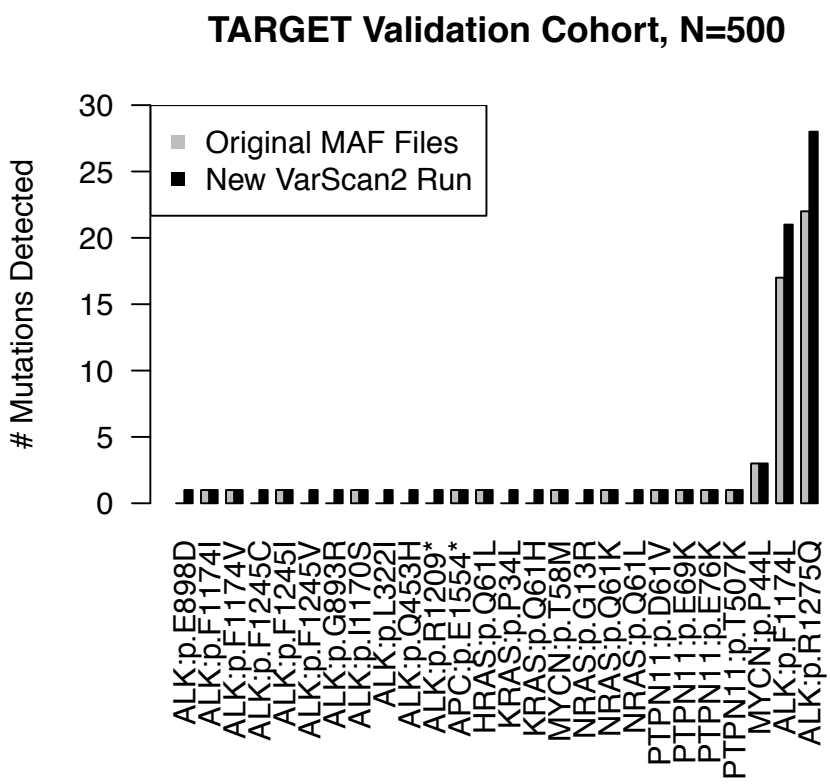

Figure 3.12: Comparison of original TARGET analysis to reanalysis of low frequency variants in 500 tumor/normal pairs. The TARGET validation sequencing dataset was reanalyzed using VarScan2 with the option "–min-var-freq 0.05" to detect low variant allele fraction (VAF) mutations. As illustrated here, a large number of driver mutations in important oncogenes such as *ALK*, *KRAS*, and *NRAS* were missed by the prior analysis, but are detected upon reanalysis.

Similar to our relapse study, we will use statistical analysis to predict whether low variant frequency is primarily the result of subclonality versus low sample purity. These efforts will have important clinical significance in predicting the proportion of patients who might benefit from targeted therapies, and may aid in the establishment of guidelines for minimum depth of coverage for detecting driver mutations from targeted sequencing.

Chapter 4

Toward new treatments for neuroblastoma

4.1 Targeting the GATA2/3-LMO1 signaling axis

Our discovery that rs2168101 modulates *LMO1* gene expression *in cis* by disrupting a GATA transcription factor binding site has uncovered a GATA2/3-LMO1 signaling axis in neuroblastoma that can perhaps be exploited therapeutically. Although transcription factors are often thought of as "undruggable" by traditional small molecule inhibition, their central role in oncogenic signaling makes them attractive targets for new therapeutic strategies, which include inhibition of transcription factor expression (e.g. with siRNA or miRNA), inhibition of DNA binding (e.g. with oligodeoxynucleotide decoys or pyrrole-imidazole polyamides), and epigenetic modulation of transcription factor function (e.g. BET bromodomain inhibitors)[128]. One relevant and timely example is SB010, a novel anti-GATA3 DNA enzyme therapeutic that has shown promising results for the treatment of allergic asthma[129]. However, further study is warranted in order to determine if SB010 may show therapeutic efficacy in neuroblastoma and to what extent GATA3 inhibition efficacy may depend on rs2168101 genotype—i.e. to what extent the oncogenic effects of GATA3 are dependent on downstream LMO1 signaling.

Inhibiting LMO1 directly or targeting its downstream effectors may present another and potentially more viable therapeutic opportunity. In this regard, our ongoing efforts to map the downstream targets of LMO1 by LMO1 ChIP-seq are crucial, as these may uncover additional effectors of oncogenic signaling that are more amenable to small molecule inhibition or other therapeutic modalities. Furthermore, in addition to the broad strategies outlined above for targeting GATA3, it may be possible to affect LMO1 signaling through inhibition of protein-protein interactions (e.g. via peptide aptamer or antibody fragments), a strategy which has shown efficacy when targeted against the better studied LMO1 paralog, LMO2, in preclinical models[130].

Given that GATA2/3 and LMO1 play important roles in normal cellular processes in addition to promoting oncogenesis in neuroblastoma, it is unclear if on-target, off-tumor toxicities may limit the therapeutic potential of systemic inhibition of LMO1 signaling. Antibody-directed therapy including drug conjugates and new immunotherapy targeted to neuroblastoma cell surface molecules—including ALK, GD2, and NCAM, among others—may provide a method to circumvent such toxicities, and further studies are required to ascertain if and how the cell surface repertoire of LMO1-driven neuroblastomas differs in comparison to other forms of neuroblastoma. On the other hand, the recent and rapid development of CRISPR-Cas9 gene editing technologies[131–133] opens the possibility of targeted enhancer element editing both *in vitro* and *in vivo*, which could be utilized to directly ablate the activity of the tissue-specific *LMO1* enhancer investigated here. However, as concerns rise over the lack of specificity and off-target editing inherent to the CRISPR-Cas9 system[134], the development of novel nuclease dead CRISPR-Cas9 proteins fused to chromatin remodelers[135, 136] may provide another therapeutic opportunity to ablate *LMO1* enhancer activity through targeted editing of the epigenome without the inherent risks associated with DNA-editing.

4.2 New therapies for relapsed neuroblastoma

Our identification of RAS pathway activation in a substantial fraction of neuroblastomas that relapsed on chemotherapy indicates that targeting the RAS pathway may provide a route to rational therapy when more traditional treatments fail. Indeed, the RAS pathway has long been viewed as an attractive target for new therapies due the high prevalence of RAS-activating mutations in cancer more broadly: approximately one-third of solid tumors and one-fifth of myeloid malignancies[137]. Much of the biochemistry of RAS-signaling has been worked out, and it is now understood that RAS proteins are guanine binding proteins (G proteins) where dynamic cycling between GTP- and GDP-bound states regulates cellular proliferation in normal physiology; to summarize, the GTP-bound state induces conformational changes in RAS that activate additional downstream growth signaling effectors whereas GDP-binding is suppressive[137]. And yet despite these biochemical insights and over three decades of intense research, effective rational therapies targeting the RAS pathway have largely eluded the cancer community. Attempts to drug the GTP-binding site of RAS have proven considerably less fruitful than inhibiting tyrosine kinases via their ATP-binding domains, due to the comparatively high picomolar affinity for GTP, the high intracellular abundance of GTP, and the lack of apparent druggable allosteric sites[137].

Due to the difficulty of inhibiting RAS directly, many alternative strategies are being explored to treat RAS-mutated cancers, including targeting downstream RAS effectors, targeting RAS localization, targeting synthetic lethal partners of RAS, and immune-based therapies, though none of these have yet yielded robust clinical responses[137]. Indeed, our studies of single-agent MEK inhibition in RAS-pathway activated neuroblastoma cell line models elicit cytostatic responses that delay but

do not prevent eventual tumor outgrowth (Figure 3.10). Our laboratory is therefore actively investigating combination therapies, and we have performed synergy screens in order to identify drug combinations that exhibit superadditive anti-tumor effects in RAS pathway activated neuroblastoma. These efforts identified combined ALK and CDK4/6 inhibition (ceritinib + ribociclib) as a potent combination for ALK-mutated neuroblastomas, as well as combined MEK and CDK4/6 inhibition (trametinib + ribociclib) as a potent combination for neuroblastomas that are *MYCN*-amplified and/or harbor other non-ALK lesions in the RAS-MAPK pathway. Remarkably, CDK4 was recently identified as a synthetic lethal partner of KRAS in non-small cell lung cancer[138] and was also identified as the principal driver of the molecular differences between the genetic and pharmacological perturbation of NRAS in melanoma[139], suggesting an integral link between CDK4 signaling and RAS signaling more broadly. In addition to our discovery of a high prevalence of RAS-MAPK activating mutations in relapsed neuroblastomas, these results provide the rationale for using exactly these combinations in patients with appropriately matched genomic lesions in the soon to open NEPENTHE phase 1b/2 clinical trial.

Although these data give us cautious optimism that we can prolong survival for relapsed neuroblastoma, the lack of permanent responses to combined MEK and CDK4/6 inhibition in our preclinical studies suggests that new treatment strategies will likely be required for curative therapy in patients with RAS-MAPK activation in the absence of ALK mutations. In the relative short term, higher order drug combinations can be investigated for their potential to induce even better clinical responses. Longer term, the development of entirely new treatment approaches have the potential to revolutionize the field of RAS-targeted therapy, including new allosteric inhibitors that directly target mutated RAS (for example, a novel inhibitor that covalently and allosterically modifies mutant KRAS G12C to preferentially bind inactivating

GDP[140]) and the emergence of new immunotherapies (for example, an anti-GD2 therapy recently approved for high-risk neuroblastoma[8]). Indeed, targeting RAS has recently experienced a resurgence in interest both in academia and industry, culminating in the launch of NCI RAS Initiative in 2013, with the hope of uncovering new and transformative therapies[141].

In the immediate term, basic questions still remain concerning how RAS pathway activating mutations arise and evolve clonally in neuroblastoma, which are likely to have therapeutic implications. The higher prevalence of RAS-MAPK mutations in relapsed neuroblastoma observed in our study[106] in comparison to our German colleagues[125] may originate from much higher rates of chemotherapy treatment in our cohort or other biases in sample collection. It will therefore be important to better ascertain the prevalence of RAS-MAPK activating mutations prospectively through efforts such as our NEPENTHE trial, and sequencing could be more broadly incorporated into our pre-clinical treatment models (e.g. cell line or patient derived xenografts) to better understand the molecular correlates of drug resistance and how they evolve over the course of treatment. Rapid progress has been made in developing single cell DNA and RNA sequencing technologies in recent years[142], which could be applied to achieve an even finer resolution view of the clonal evolution events that underlie neuroblastoma relapse and treatment resistance. In light of our observation of RAS-MAPK activating mutations at relapse that are undetectable even by ultra-deep sequencing in matched primary tumor biopsies, it is an open question whether or not adding ALK and/or MEK and/or CKD4/6 inhibition as frontline therapy could be used to anticipate and/or prevent the outgrowth of RAS-MAPK activated subclones when they are undetectable at diagnosis. All of these questions provide interesting avenues for future studies.

4.3 Final remarks

In conclusion, the field of neuroblastoma genomics has experienced rapid progress in just the past 5-10 years, with key advances including the discovery of causal mutations for familial neuroblastoma, discovery of common variation associated with sporadic disease, mapping the somatic coding landscape of neuroblastoma primary tumors, and the emergence of mutated *ALK* as a therapeutic target for rational therapy. However, much work remains to be done to understand how germline and somatic lesions interact to promote tumor initiation and maintenance, to elucidate the epigenetic landscape of disease subtypes, to better define the temporal evolution of neuroblastoma during tumor initiation and under different treatment modalities, and to ultimately develop new therapeutic strategies for the majority of neuroblastomas which lack traditionally "targetable" mutations and for relapsed disease.

The studies that constitute this dissertation begin to address some of these important questions and provide a blueprint for progress in the field. The identification of a causal high-risk-associated germline variant in a *LMO1* super-enhancer has laid the groundwork for comprehensive epigenomic profiling efforts to understand the regulatory landscape of neuroblastoma, generating data that is likely to reveal insights into LMO1 function specifically, genome-scale regulatory architecture more broadly, and provide insights into the function of non-coding mutations. The discovery of clonal evolution of RAS-MAPK mutations in relapsed neuroblastoma is now directing efforts to better define the subclonal makeup of neuroblastoma tumors as well as an early phase clinical trial of MEK-inhibitors in relapsed neuroblastoma. In the coming years, we are hopeful that these studies will be looked back upon as crucial steps toward a comprehensive understanding of the molecular basis of neuroblastoma and the development of better treatments for this aggressive childhood malignancy.

Appendix A

Statistical methods for estimating subclonality of point mutations from NGS data

A.1 ABSOLUTE: Carter *et al.* 2012

Published in 2012, ABSOLUTE was a first-in-class algorithm developed to aid in the estimation of clonal versus subclonal mutations detected by NGS experiments from cancer samples [107]. Because tumor purity and ploidy are important confounders in the inference of mutation clonality, the primary goal of ABSOLUTE is to jointly infer the purity and ploidy from NGS data and to discern clonal versus subclonal events. The theoretical framework of ABSOLUTE is summarized in what follows.

For simplicity, we consider a mixed population which includes a fraction α of a clonal cancer cell population and a fraction $(1 - \alpha)$ of contaminating normal cells; in the idealized case the normal cells will be diploid (copy number = 2) across the whole genome. Let $q(x)$ denote the absolute integer copy number at locus x in the cancer cells. The average tumor "ploidy", τ, is therefore the absolute copy number averaged

over the N segmented loci across the whole genome:

$$\tau = \sum_{x=1}^{N} w(x)q(x)$$

where the weight of each locus, $w(x)$ scales proportionally to the size of the locus in base pairs. The relative copy number, R, of locus x can be expressed as:

$$R(x) = \frac{2(1-\alpha) + q(x)\alpha}{2(1-\alpha) + \tau\alpha}$$

For a somatic point mutation (e.g. SNV, indel, or SV breakpoint) at locus x, we can express the expected fraction of mutant alleles, F, as follows:

$$F(x) = \frac{s(x)\alpha}{2(1-\alpha) + [q(x) - s(x)]\alpha + s(x)\alpha} = \frac{s(x)\alpha}{2(1-\alpha) + q(x)\alpha}$$

where $s(x)$ represents the absolute copy number of the mutation (aka "somatic multiplicity"), $2(1-\alpha)$ represents the relative weight of DNA originating from normal cells, $[q(x) - s(x)]\alpha$ represents the relative weight of non-mutant DNA originating from cancer cells, and $s(x)\alpha$ represents the relative weight of mutated DNA originating from cancer cells. Note that $1 \leq s(x) \leq q(x)$ generally, but that $s(x) < q(x)$ will hold assuming that the mutation only arose once in the evolutionary history of the cancer and that LOH has not occurred. Practically speaking, the quantities $R(x)$ (relative copy number) and $F(x)$ (mutant allele fraction) can be easily and directly estimated by NGS, whereas $q(x)$ and $s(x)$ are the quantities of biological interest that are less straightforward to estimate, with α and τ as nuisance parameters.

Within this framework, $R(x) = \frac{2(1-\alpha)}{2(1-\alpha)+\tau\alpha}$ for clonally homozygously deleted sites—i.e. $q(x) = 0$. For adjacent copy-number states – i.e. for each unit increase in $q(x)$ – it follows that $R(x)$ will correspondingly increase by $\frac{\alpha}{2(1-\alpha)+\tau\alpha}$. Outliers from this

pattern are inferred to be "subclonal" CNAs. At a fixed $q(x)$, a similar pattern exists for clonal SNVs – i.e. $F(x) = \frac{s(x)\alpha}{2(1-\alpha)+q(x)\alpha}$ for $s(x) \in \{1, 2, ..., q(x)\}$ – and outliers from this pattern are inferred to be "subclonal" SNVs.

A.2 Extending ABSOLUTE: Landau *et al.* 2013

In Landau *et al.* 2013, this framework was extended to estimate the cancer cell fraction (CCF) of SNVs (CCF is the fraction of tumor cells harboring a presumed subclonal mutation) rather than simply call presence of such mutations as outliers [108]. Their method specifically applies to subclonal SNVs that may arise within a non-CNA or clonal CNA region, resulting in unit somatic multiplicity ($s(x) = 1$) under the parsimonious assumption that the SNV mutation only arose once within the evolutionary history of a specified tumor (Figure A.1). As a function of the CCF, c, the expected fraction of mutant alleles will therefore follow the equation:

$$F(c) = \frac{\alpha c}{2(1-\alpha) + q(x)\alpha}$$

Thus, for a particular SNV with a specified CCF equal to c, the likelihood of observing a mutant alleles out of n total sequencing reads equals

$$P[A = a | C = c, N = n] = \text{Binomial}(a; n, p = F(c))$$

so that

$$P[C = c | A = a, N = n] \propto \text{Binomial}(a; n, p = F(c))$$

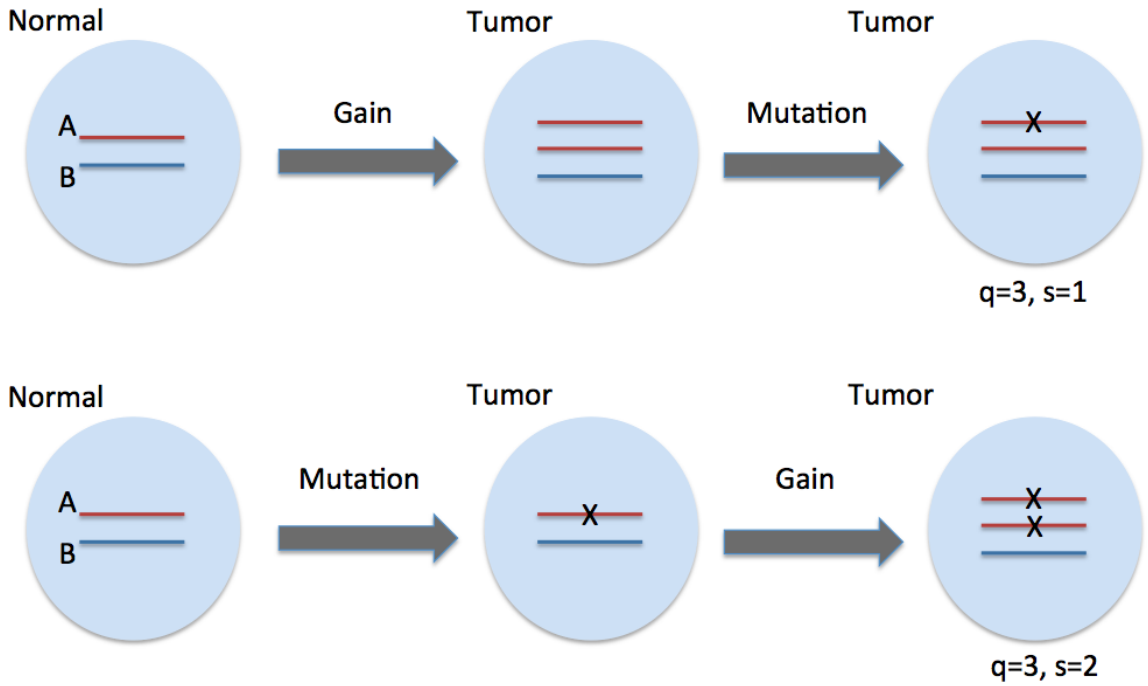

Figure A.1: Two examples of sequential tumor evolution. (Top) In this example, copy-number gain precedes somatic point mutation at a given locus x, resulting in copy-number $q(x) = 3$ and somatic multiplicity $s(x) = 1$. (Bottom) In this example, somatic point mutation is followed by sequential gain of the mutated allele at a given locus x, resulting in copy-number $q(x) = 3$ and somatic multiplicity $s(x) = 2$.

will hold under a uniform prior on C[1]. In practice, α and $q(x)$ are estimated by ABSOLUTE or related algorithms (our laboratory prefers the Sequenza algorithm [109], due to ease of use) whereas a and n are measured directly from NGS data. The distribution of C is then computed using a single variable grid approach.

[1]Equivalently, a uniform prior implies that $F(C)$ follows a Beta$(a+1, n-a+1)$ distribution that is right-truncated at $F(1)$ due to the $0 \leq c \leq 1$ constraint. Therefore, the conjugate prior for $F(C)$ will be another Beta(a', b') distribution that is truncated at $F(1)$, with a corresponding truncated Beta$(a+a'+1, n+b'-a+1)$ posterior. Thus, a' and b' can be interpreted as "pseudocounts" for mutant and wild-type reads, respectively. While we consider $F(C)$ here primarily for mathematical convenience, note that F is a simple scaling function with respect to C, so that the posterior and conjugate-priors for $C = F^{-1}(F(C))$ are simply re-scaled versions of the distributions just described.

Appendix B

Methods for Oldridge, Wood *et al.* 2015

Genotype imputation and association testing

A primary European-American cohort of 2,101 cases and 4,202 matched controls were assayed with Illumina HumanHap550 v1, Illumina HumanHap550 v3, and Illumina Human610 SNP arrays as previously described[51]. Genotypes were phased using SHAPEIT v2.r790 and data from 1000 Genomes phase 1 version 3. Subsequently, imputation was performed using IMPUTE2 v2.3.1 for all SNPs and indel variants annotated in the 1000 Genomes phase 1 version 3. Testing for association with neuroblastoma under an additive genetic effect model was performed using the frequentist likelihood score method implemented in SNPTEST v2.4.1. Genotypes for a previously described African-American cohort of 365 cases 2491 controls[143] were imputed and tested for neuroblastoma association using the same analytic pipeline. Statistical adjustment for gender was performed in both cohorts. For population stratification adjustment, the first 20 multidimensional scaling (MDS) components were included as covariates in the European-American cohort, while a measure of African admixture as estimated by the ADMIXTURE software program was used in the African-American cohort. Manhattan plots of SNP position and statistical significance were generated using LocusZoom software. Linkage plots were generated by Haploview software based on

HapMap CEU individuals (version 3, release 2) using default settings. All research subjects or their guardians provided informed consent for research, and all institutions involved in this research had regulatory approval for human subjects research.

Prioritization of candidate causal variants

All SNPs and indels reported in the 1000 Genomes phase 1 version 3 data were considered as candidate causal variants and were ranked based on a combination of (1) neuroblastoma association in the primary European-American cohort, (2) evolutionary conservation, (3) DNase I hypersensitivity, and (4) transcription factor binding motif matching. Neuroblastoma association in European-Americans was evaluated as described above. Conservation scores were computed as the average of the phastCons46wayPlacental UCSC conservation track score for all bases from the -10 position to the $+10$ position surrounding each candidate variant. A DNase I hypersensitivity score was calculated by counting the number of sequencing tags from the -100 position to the $+100$ position around each candidate variant in ENCODE data for the neuroblastoma cell line, SK-N-SH. Position weight matrices representing transcription factor binding motifs were obtained from the JASPAR database, and candidate binding sites were identified by scanning the hg19 human reference genome using the MATCH-TM algorithm with a matrix similarity score (mSS) threshold of 0.90.

Neuroblastoma association replication and meta-analysis for rs2168101

We replicated the association of rs2168101 with neuroblastoma by direct genotyping of rs2168101 in independent Italian (cases = 420, controls = 751) and UK cohorts (cases = 369, controls = 1109). Meta-analysis across the European-American, African-American, Italian and UK cohorts was performed using the inverse variance method provided in the METAL software program. Beta values (log-odds) and standard errors generated by SNPTEST, as described above, were used as input.

Survival analysis

We compared both overall survival and event-free survival over a 10-year follow-up period between G/G versus G/T and T/T rs2168101 genotypes in a case-case comparison between neuroblastoma patients from the European-American cohort. Because rs2168101 genotypes were imputed in this cohort, the most-probable genotype predicted by IMPUTE2 was used for each patient. In the event of insufficient follow-up, all data was right censored. Cox proportional hazard modelling was performed using 20 MDS components to account for population stratification, in addition to MYCN amplification status, as covariates. All statistical analysis and generation of Kaplan-Meier plots was performed in R using the CRAN repository package, "survival".

Total and allele-specific expression analysis

Total and allele-specific RNA expression analysis was performed based on poly-A-enriched RNA-sequencing data from 127 primary neuroblastoma tumours sequenced through the TARGET project. RNA-seq reads were aligned to the hg19 human reference genome using the STAR aligner (v2.4.0b). Aligned reads were assigned to RefSeq genes using HTSeq (v0.6.1) and normalized to RPKM for total gene expression measurements. DNA genotypes for rs2168101 were obtained either through matched whole-genome sequencing ($n = 69$) or targeted genotyping assays ($n = 58$ additional tumours). DNA genotypes for rs3750952 were obtained through either matched whole-genome or whole-exome sequencing.

Allele-specific RNA expression analysis was performed from a subset of 45 primary neuroblastoma tumours (out of 127) with the necessary synonymous exonic SNP genotypes (rs3750952 = C/G) to enable measurement of allelic expression by mRNA-seq. As a readout for allelic imbalance of rs3750952, we computed allelic fractions as min(C, G)/(C + G), since phasing between rs3750952 and rs2168101 alleles in each tumour was unknown. Statistical comparison between the two groups was performed

by two-sided Welch's t-test, comparing 12 tumours heterozygous for rs2168101 (G/T) to the remaining 33 tumours that were homozygous for rs2168101 (G/G) as controls. DNA genotyping for rs2168101 was performed by whole-genome sequencing or a directed genotyping assay, whereas DNA genotyping for rs3750952 was determined from TARGET whole-exome or whole-genome sequencing. Where possible, integrity of sample matching was verified by measurement of genome-wide genotype concordance. All genotypes are reported with respect to the minus strand of the human reference genome, hg19.

To measure allele-specific expression directly at the intronic SNP we first purified the nuclear RNA fraction using the Cytoplasmic and Nuclear RNA purification Kit (Norgen Biotek, 21000) from four neuroblastoma cell lines (SNP rs2168101: SHSY5Y = G/G; NLF = G/T; NGP = G/G; NB1643 = T/T). Ion AmpliSeq Designer v3.4.3 (Life Technologies White Glove service) was used to design amplicons targeting the intronic SNP rs2168101 and three additional exonic SNPs in linkage disequilibrium. Custom AmpliSeq libraries were prepared in triplicate for each cell line, indexed, pooled and sequenced using an Ion 318 Chip on a Personal Genome Machine (Life Technologies). Reads were aligned to the hg19 reference genome and a synthetic genome showing the alternate allele at SNP rs2168101 at hg19 chr11:8255408 to account for any alignment bias. High-quality mapped reads containing the reference G allele or alternative T allele were counted and tested for significant deviation from 50:50 expression using a two-sided one-sample t-test (null hypothesis that allele fraction = 0.50) across three experimental replicates. Primer pair sequences:

For rs1042359:

forward: 5'-GTGTGGGAGACAAAUTCTTCCUGA-3',

reverse: 5'-GCCGGGCGUTACTGAACUT-3';

For rs3750952:

forward: 5'-CGCAAGAUCAAGGACCGCTAUC-3',

reverse: 5'-GATGAGGTUGGCCTTGGTGUA-3';

For rs2168101:

forward: 5'-CCUTTCCUGAAGGAGCGCAAA-3',

reverse: 5'-CACTTTCCATUAAGGAGATAGCAUCCC-3';

For rs204929:

forward: 5'-CAAUCTAGGTUAAGAGCCGGACAAG-3',

reverse: 5'-GTGUCCAGCCGCAGCUA-3'.

Reporter assays

Primers were designed to clone a 553-bp genomic region (hg19, chr11:8255155-8255707) surrounding the candidate SNP rs2168101 at the GATA transcription factor binding site from neuroblastoma cell lines SKNSH (G/G) and matching site of BE2C (T/-). The cloned region did not contain other statistically significant SNPs at the LMO1 locus. The primers were designed to introduce sequences for restriction sites 5'-XhoI and 3'-BglII, which are present in the MCS of pGL4.26[luc2/minP/Hygro] (Promega, E8441). XhoI/BglII restriction enzyme digested fragments were sequence verified, gel purified, ligated into pGL4.26[luc2/minP/Hygro], transformed into One Shot TOP 10 chemically competent cells (Life Technologies, C4040-10) and grown on LB plates containing 50 μg ml-1 ampicillin overnight at 37 °C. Colonies positive for the vector containing the insert were grown in 50 ml LB broth containing 50 μg ml-1 ampicillin and plasmids were purified using a Qiagen Plasmid Midi Prep Kit (Qiagen, 12143). Transfection into HEK293 cells which were approximately 50% confluent was accomplished using Fugene 6 Transfection reagent (Promega E2691) at a 3 μl:1 μg fugene:DNA ratio. Cells underwent selection in 150 μg ml-1 Hygromycin B (Mediatech, 30-240-CR) and individual colonies were picked and grown, and genotypes of constructs were confirmed by fragment size and Sanger sequencing. Subsequently,

HEK293 + 553 bp insert cells and HEK294 + vector only cells were grown in 96-well optical plates. On day 2, the cells were transiently fugene transfected with the Renilla expression control vector pGL4.74[hRLuc/TK] (Promega, E6921) at a 1:500 dilution with respect to the luciferase vector. Luciferase assays were carried out 48 h after Renilla transfection using Dual Luciferase Reporter Assay System (Promega, E1910) with read-outs performed on a Dual Injector System for GloMax-Multi Detection System (Promega, E7081). Luciferase expression was normalized to Renilla expression. All reporter assays were performed in quintuplicate (five technical replicates each) across the experimental conditions: (1) HEK293T, (2) HEK293T with empty vector, (3)-(6) four independent clones of HEK293T with T-allele construct, and (7)-(10) four independent clones of HEK293T with G allele construct. Results were averaged across technical replicates, normalized to empty vector, and reporter activities for T allele versus G allele clones (four biological replicates each) were analysed by two-sided Welch's *t*-test.

Construct risk allele (G):

GTAGGGGTTGGAGTTCAGCCTGTTTCCCCTCCAATGTTGTTCCCCCC
ACATCCTGAGACTTAGGGGTGACCCTGGGTTGAGTGGACTGGTTTATTC
TGCTGGGCCCAGCGCATGCATCTGAGTGTGTGCCCAGGCGTGCGTGTCG
GCGCAAACATCATCCATTGTGAAATATCAGTGTTTTCATGGGTGAGTAG
TAATTACTGGGTAATGCTTTAAAACCTTTCCTGAAGGAGCGCAAAGCCA
TTTTTTTCTAAAGTCAGGAGTACATTAAAAGGATTACCATGTAGATTTG
ATTTTTAGATAACACTAAAATGGATCCCAAATGGACTTCAGCAAAGGGA
TGCTATCTCCTTAATGGAAAGTGCATGGCCCGAGGCTCAGGTCCCAGAG
CCAGGCTGGGGAAGGAGGGAGGGAAGAGGTGTCTGCAGGGGGGCAGGC
TGGCAGATTGGGTGGGGGCTAGGTGGGAATGGGGAAGGCAGAGCAGGA
GGGAGGGCCTGGACCCTGTGGGGAGCTTATCCCTCCATCTGGGGAGCAG

GAGACTACAGAGCCCCT

Construct protective allele (T):

GTAGGGGTTGGAGTTCAGCCTGTTTCCCCTCCAATGTTGTTCCCCCC
ACATCCTGAGACTTAGGGGTGACCCTGGGTTGAGTGGACTGGTTTATTC
TGCTGGGCCCAGCGCATGCATCTGAGTGTGTGCCCAGGCGTGCGTGTCG
GCGCAAACATCATCCATTGTGAAATATCAGTGTTTTCATGGGTGAGTAG
TAATTACTGGGTAATGCTTTAAAACCTTTCCTGAAGGAGCGCAAAGCCA
TTTTTTTCTAAAGTCAGGAGTACATTAAAAGGATTACCATGTAGATTTG
ATTTTTATATAACACTAAAATGGATCCCAAATGGACTTCAGCAAAGGGA
TGCTATCTCCTTAATGGAAAGTGCATGGCCCGAGGCTCAGGTCCCAGAG
CCAGGCTGGGGAAGGAGGGAGGGAAGAGGTGTCTGCAGGGGGGCAGGC
TGGCAGATTGGGTGGGGGGCTAGGTGGGAATGGGGAAGGCAGAGCAGG
AGGGAGGGCCTGGACCCTGTGGGGAGCTTATCCCTCCATCTGGGGAGCA
GGAGACTACAGAGCCCCT

Cell culture and protein lysates

Jurkat T-ALL and neuroblastoma cell lines were sourced from the American Type Tissue Culture Collection, and kept in growth medium of RPMI+10% heat-inactivated FCS with 1% penicillin-streptomycin, as previously described[81]. Cells were lysed for protein, with subsequent protein quantified by spectrophotometry, as previously described[144]. Protein was resolved on 8-14% Tris-Bis gels, transferred to PVDF membranes, blocked and subjected to primary and secondary antibodies, as previously described[144]. Primary antibodies were anti-GATA3 (Pierce Biotechnology, 1:1,000), anti-LMO1 (Bethyl Laboratories, 1:1,000) and α-tubulin (Cell Signaling Technologies, 1:1,000). Blots were developed with secondary horseradish peroxidase (HRP)-conjugated antibodies (Cell Signaling Technologies, 1:5,000) and Protein-plus Dura ECL Reagent (Thermo-Fisher Scientific). All cell lines are genotyped semiannually

to assure identity and also tested routinely for mycoplasma contamination.

Genome-wide occupancy analysis

ChIP coupled with massively parallel DNA sequencing (ChIP-seq) was performed as previously described[145, 146]. The following antibodies were used for ChIP: anti-H3K27ac (Abcam, ab4729) and anti GATA3 (Santa Cruz, sc-22206X). For each ChIP, 10 μg of antibody was added to 3 ml of sonicated nuclear extract. Illumina sequencing, library construction and ChIP-seq analysis methods were previously described[146].

ChIP-seq processing

Reads were aligned to build hg19 of the human genome using bowtie with parameters -k 2 -m 2 -e 70 -best and -l set to the read length[147]. For visualization in the UCSC genome browser[148] in Figures 2.11a and 2.12a, WIG files were created from aligned ChIP-seq read positions using MACS 1.4.2 with parameters -w -S -space = 50 -nomodel -shiftsize = 200 to artificially extend reads to be 200 bp and to calculate their density in 50-bp bins[149]. Read counts in 50-bp bins were then normalized to the millions of mapped reads, giving reads per million values.

ChIP-seq allele specificity analysis

To determine preferential ChIP-seq coverage of one allele, which implies preferential binding of protein to one allele vs. another, we counted the reads at rs2168101 using samtools mpileup[150]. By using the aligned reads described above, this gave us a count of reads with a given base at this position. The fraction of reads with the risk allele versus the protective allele is reported in Figure 2.11b. Statistical tests for preferential allelic binding were performed by two-sided binomial test.

Enriched regions

Regions enriched in ChIP-Seq signal were identified twice using MACS with corresponding control and parameters -keep-dup = all and -p 1e-9 or -keep-dup = 1 and -p 1e-9. Super-enhancers in SHSY5Y and KELLY were identified using ROSE (`https:`

//bitbucket.org/young_computation/rose[86, 151] with modifications based on ref. [82]. In brief, peaks of H3K27ac were identified using MACS as described above and their union was used as constituent enhancers. These peaks were stitched computationally if they were within 12,500 bp of each other, although peaks fully contained within ± 2,000 bp from a RefSeq promoter were excluded from stitching. These stitched enhancers were ranked by their H3K27ac signal (length × density) with input signal in the corresponding region subtracted. Super-enhancers were separated from typical enhancers by geometrically determining the point at which the line $y = x$ is tangent to the curve of stitched enhancer rank versus stitched enhancer signal. Those stitched enhancers above this point are considered super-enhancers.

To account for the known focal amplification of the MYCN locus in KELLY, BE2, BE2C and NGP neuroblastoma cells, which contain enhancers, we modified our pipeline slightly. Because MACS is insensitive for the identification of peaks in focally amplified DNA, we identified peaks of H3K27ac versus input using MACS2 callpeak (https://pypi.python.org/pypi/MACS2) with parameters -broad -keep-dup = 1 -p 1e-9 and -broad -keep-dup = all -p 1e-9. The union of these MACS2 calls was used as constituent enhancers for ROSE with the remaining parameters as described above. For Figure 2.12b, most of the curve represents the analysis performed using MACS-identified constituents; the rank and signal of the MYCN-associated enhancer comes from this MACS2-identified set of constituents to remain consistent with the conclusions and methods as previously described[82]. The curve output from the MACS-identified enhancers was vertically compressed and a point representing the signal of the MYCN-associated super-enhancer from the MACS2-identified enhancers was added in Illustrator. Super-enhancers were assigned to the single expressed RefSeq transcript whose transcription start site was nearest the centre of the stitched region. Expressed genes were in the top 2/3 of RefSeq transcripts ranked by their

promoter (transcription start site ± 500 bp) H3K27ac signal determined by bamToGFF (`https://github.com/BradnerLab/pipeline`) with parameters -e 200 -m 1 -r -d.

Clone cell generation

LMO1 cDNA was amplified from pcDNA3-LMO1 and subcloned into the XhoI and NotI site of the lentiviral vector pOZ-FHN. Lentivirus expressing FH-LMO1 was propagated in HEK293T cells by cotransfection with psPAX2 and pMD2.G plasmids (adgene) using FUGENE 6 (Roche) by standard methodologies[152]. Viral supernatant was recovered and KELLY cells were infected with lentivirus expressing FH-LMO1 or empty vector alone, as previously described[81]. Cells were sorted for expression of the IL2R, and positive expression was used to establish single cell clones. Expression of FH-LMO1 was assessed by western blotting as above to confirm overexpression.

siRNA and growth assays

SHSY5Y, KELLY and KELLY clone cells were reverse transfected with 100 nM concentrations of either non-targeted (control siRNA-1) or GATA3-targetted siRNA-1 or -2 (Ambion) for 6 h with lipofectamine 2000 (1:1,000) in Optimem I before being replated into growth assays in normal RPMI growth media. Cells (2×10^5) were replated in triplicate for counting at 24, 48 and 72 h post-transfection by manual hemocytometry. Cells (5×10^5) were replated for protein lysates at the same time points. All experiments were repeated in triplicate, with a technical replicate number of 9 for all cell growth assays as described[153]. Statistical tests were performed by two-sided Welch's t-test.

Data access

GWAS and sequencing data used for this analysis are available in dbGaP under accession phs000124 and phs000467. The tumour genomics data are also available through the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) data matrix portal (`http://target.nci.nih.gov/dataMatrix/TARGET_DataMatrix.`

html). Data generated through the ENCODE project including DNase I hypersensitivity sequencing and ChIP-sequencing data were obtained from ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/. Aligned sequencing read (bam) files were used as provided from the FTP site. The mammalian evolutionary conservation track representing 46 mammalian species (phastCons46wayPlacental) was obtained from the UCSC Table Browser http://genome.ucsc.edu/cgi-bin/hgTables?command=start. JASPAR-annotated transcription factor binding site position frequency matrices were obtained from http://jaspar.genereg.net/html/DOWNLOAD/JASPAR_CORE/pfm/nonredundant/pfm_all.txt. New ChIP-seq data sets generated in this study are available under super series GSE65664.

Appendix C

Methods for Eleveld, Oldridge, Bernard *et al.* 2015

Sample collection and patient selection.

The inclusion criteria for this study were histopathological confirmation of neuroblastoma at original diagnosis and the presence of biopsy material from a subsequent relapse specimen. Patients were included in this study after obtaining informed consent from parents or guardians, with oversight from the ethics committees 'Comité de Protection des Personnes Sud-Est IV', reference L07-95/L12-171, and 'Comité de Protection des Personnes Ile-de-France', reference 0811728 in France, the review board at the Children's Hospital of Philadelphia and review boards at other Children's Oncology Group sites that submitted samples for patients on this study in the United States, and the review board of the Academic Medical Center of the University of Amsterdam in the Netherlands. Somatic *ALK* mutation status has been reported for samples FR_NB0175, FR_NB308, FR_NB399, FR_NB1224, FR_NB1269 and FR_NB13821[104].

Whole genome sequencing, summary.

Whole-genome sequencing was performed by Complete Genomics to an average coverage of 50x per sample (for Dutch and US patient material and cell line material)[154] or using Illumina HiSeq 2500 instruments to an average coverage of 80x per sample (for French patient material). Material for each patient (lymphocytes, primary tumor and

relapse tumor) was in all cases sequenced using the same sequencing platform. Estimation of normal tissue contamination and whole-genome segmentation was performed on all primary-normal and relapse-normal pairs with Sequenza software v2.1.0 using both binned coverage ratio data and SNV allelic ratios as input[109]. Coverage-based copy number plots were generated as previously described[44], with the exception that values were corrected for ploidy and purity as determined by Sequenza. The R2 bioinformatics platform and Circos[155] were used for analysis and visualization.

Whole genome sequencing, Complete Genomics.

Potential somatic variants were determined with the CallDiff algorithm with somatic output as available in the CGAtools v1.3.0 package, maintained by Complete Genomics. Every tumor was compared to its matched blood sample across the genome. The somatic output files were then filtered to regions where coding sequences are defined in the UCSC refflat. Silent mutations were subsequently removed from the analysis. In addition, we determined the presence of somatic splice-site variants in the three bases surrounding exons as defined by UCSC refflat data. Variants with a somatic score > 0.05 (for NL_041, NL_571 and NL_N607) or Somatic Quality high (the remainder of the patients) were included in the analysis.

Comparisons of structural variants between tumor and lymphocyte genomes were performed with the JunctionDiff and Junction2Event algorithms from CGAtools. These somatic events were filtered to remove events matching the following criteria: events annotated as artifacts, footprints smaller than 70 bases, less than 10 discordant mate pairs, under-represented repeats and presence in a set of baseline genomes (as provided on the website of Complete Genomics (B36baseline-junctions.tsv)). Of the remaining entries, we kept the following events: (i) exon_bites, where both ends of a junction were within the same gene and in addition affect an exonic sequence, (ii) breaks by inversion, where both ends of a junction land within a gene, thereby damaging

both genes but leaving the genes in between unaffected, (iii) potential fusion genes, which are strand matched, where both ends of a junction landed within a gene and the resulting end product fit in terms of orientation of both genes, and (iv) regions (deletions or (tandem) duplications) of up to 1 Mb, containing up to five genes.

Whole genome sequencing, Illumina.

Whole-genome sequencing was performed using an Illumina HiSeq 2500, with 90-bp paired-end reads for six tumors and 100-bp paired-end reads for the remaining two tumors. After alignment with hg19 using Burrows-Wheeler Aligner (BWA)[156], bam files were cleaned according to Genome Analysis Toolkit (GATK) recommendations[157].

Variant calling was performed in parallel using three variant callers: GATK 2.2-16, SAMtools 0.1.18 and MuTect 1.1.430 (refs. [150, 157, 158]). ANNOVAR v2012-10-23 (ref. [159]) with COSMIC v64 and dbSNP v137 was used for annotation. SNVs with a quality under 30, a depth of coverage under 6 or with fewer than 2 reads supporting the variant were filtered out, as were variants reported in more than 1% of the population in the 1000 Genomes Project[160] or Exome Sequencing Project (Exome Variant Server, National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project (ESP)).

Variants were then filtered to those regions where coding sequences are defined or to variants in the three bases surrounding exons. Silent mutations were subsequently removed from the analysis. Tumor and corresponding constitutional genomes were compared using the SAMtools mpileup algorithm[150], and non-somatic variants were discarded from the analysis.

Structural variants, including deletions, inversions, tandem duplications and translocations, were analyzed using DELLY v0.5.5 with standard parameters[161]. In tumors, at least ten supporting reads were required to make a call, and five supporting reads were required for sample NB0175 with a coverage of only 40x. To predict structural

variants in constitutional samples for subsequent somatic filtering, only two supporting reads were required. To identify somatic events, all the structural variants in each normal sample were first flanked by 500 bp in both directions and any structural variant called in a tumor sample that was in the combined flanking regions of the respective normal sample was removed. Deletions with more than five genes affected or larger than 1 Mb in size and inversions or tandem duplications covering more than four genes were removed. We focused on exonic and splicing events for deletions, inversions and tandem duplications. For translocations, we kept all structural variants that occurred in intronic, exonic, 5' UTR, upstream or splicing regions.

Clonality analysis.

Estimation of the cancer cell frequency of somatic mutations was performed using the Bayesian method of Carter *et al.* to infer posterior intervals without clustering for comparison[107]. Namely, we assumed that the expected allele fraction of a mutation in a sample of tumor purity "α", total somatic copy number q, and mutation multiplicity s can be expressed as a function of the cancer cell fraction c: $f(c) = \alpha cs/(2(1-\alpha)+\alpha q)$. Given a uniform prior on c, the posterior density of c is therefore proportional to $\text{Binom}(a|N, f(c))$, where a is the variant read count and N is the total read count. While α and q are estimated by Sequenza, the mutation multiplicity s is generally not known. However, under the parsimony assumption that a mutation occurs only once within a tumor's evolutionary history, we can bound s by $1 \leq s \leq m \leq q$, where m is the major allelic copy number of the mutation locus, which is estimated by Sequenza. We therefore modeled the posterior distribution under two assumptions: $s = 1$ (biased toward higher clonality estimates) as well as $s = m$ (biased toward lower clonality estimates).

Estimation of CCF was also performed using PyClone v0.12.7 as an alternative method for comparison[162]. For each primary or relapse tumor, PyClone was run on

all somatic coding point mutations using the "parental_copy_number" method and "pyclone_beta_binomial" density, with estimates of tumor purity and allelic copy number from Sequenza provided as necessary inputs. The Markov Chain Monte Carlo (MCMC) step of PyClone was run for 10,000 iterations with burn-in and thin parameters set to 1,000 and 10, respectively, resulting in 900 independent samples from the posterior distribution of cancer cell fraction per mutation. Otherwise, default options for PyClone were used.

Cancer mutation analysis.

To identify pathways or processes that were frequently affected in neuroblastoma relapse tumors we used the CancerMutationAnalysis R package[110]. Somatic mutations detected only in the relapse and detected in the relapse and primary for all tumors were used as input. This algorithm is not suitable for the analysis of structural variants, so these were not included in this analysis. P-values were generated using the permutation null method without heterogeneity and signify enrichment of mutated genes associated with a certain Gene Ontoloty (GO) Biological Process category across all relapse tumors.

Cell lines.

All cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS, 20 mmol/L L-glutamine, 10 U/mL penicillin, and 10 μg/mL streptomycin and maintained at 37°C under 5% CO_2. Cell line identities are regularly confirmed by short tandem repeat profiling using the PowerPlex16 system and GeneMapper software (Promega). Cell lines are regularly screened for the presence of mycoplasma.

Cell viability assays in response to MEK inhibition.

For cell viability assays, 2.5-25 $\times$ 10^3 cells were seeded in 50 μl in 96-well plates 1 d before treatment with one of three MEK inhibitors. Binimetinib, trametinib or

cobimetinib (Selleckchem) was added in a seven-point fivefold dilution series, and cell viability was assayed by MTT assay (Sigma) after 72 h, as described previously[163]. All experiments were performed in triplicate, and values were compared to those for solvent-treated controls. GI_{50} values were calculated by $100 \times (T - T_0)/(C - T_0)$ for every drug concentration, where T is the optical density for a certain drug concentration at 72 h, T_0 is the optical density at 0 h (before adding the drug) and C is the optical density of solvent-treated controls at 72 h. Curves were fitted on the data points using nonlinear regression in GraphPad 5 (log(inhibitor) versus response - variable slope), and GI_{50} values were interpolated from these curves. If curves did not reach 50% growth inhibition, the GI_{50} value was set at 10 μM.

Whole exome sequencing of neuroblastoma cell lines

Whole exome sequencing of the neuroblastoma cell lines SK-N-AS, NBL-S, Kelly and IMR-5 was performed using in-solution hybrid capture[164] followed by Illumina sequencing, as described previously[46]. Cell lines were obtained from the Neuroblastoma Cell Line Repository at the Children's Hospital of Philadelphia.

MEK inhibition in murine xenotransplants

The human neuroblastoma-derived cell lines SK-N-AS, NBL-S, Kelly and IMR-5 were xenotransplanted subcutaneously into female C.B-17 SCID mice at 5-7 weeks of age. Once the engrafted tumors reached 200 mm^3, mice were treated orally with binimetinib (Novartis) at 3 mg/kg ($n = 10$) or 30 mg/kg ($n = 10$) or received vehicle only ($n = 10$) by simple randomization. Mice were treated twice daily, and tumor size was monitored three times weekly; all investigators other than the mouse technician were blinded to group allocation and study outcomes until all mice completed a trial. Tumor burden was determined according to the formula $(\pi/6)d^3$, where d represents the mean tumor diameter obtained by caliper measurement. Statistical analysis was performed using a two-tailed t-test at each time point, with P-values < 0.05 indicating

significance between the vehicle-treated group and each of the treatment groups. All studies were performed in accordance with the Children's Hospital of Philadelphia Institutional Animal Care and Use Committee, and mice were euthanized as soon as tumor volume exceeded 3 cm^3. The sample size of ten mice was predetermined on the basis of statistical power calculations.

Bibliography

[1] J. R. Park, R. Bagatell, W. B. London, J. M. Maris, S. L. Cohn, K. M. Mattay, M. Hogarty, and C. N. Committee, "Children's oncology group's 2013 blueprint for research: neuroblastoma.," *Pediatric blood & cancer*, vol. 60, no. 6, pp. 985–993, 2013.

[2] G. M. Brodeur and R. Bagatell, "Mechanisms of neuroblastoma regression.," *Nature reviews. Clinical oncology*, vol. 11, no. 12, pp. 704–713, 2014.

[3] A. Oberthuer, F. Berthold, P. Warnat, B. Hero, Y. Kahlert, R. Spitz, K. Ernestus, R. König, S. Haas, R. Eils, M. Schwab, B. Brors, F. Westermann, and M. Fischer, "Customized oligonucleotide microarray gene expression-based classification of neuroblastoma patients outperforms current clinical risk stratification.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 24, no. 31, pp. 5070–5078, 2006.

[4] Y. P. Mosse, S. J. Diskin, N. Wasserman, K. Rinaldi, E. F. Attiyeh, K. Cole, J. Jagannathan, K. Bhambhani, C. Winter, and J. M. Maris, "Neuroblastomas have distinct genomic DNA profiles that predict clinical phenotype and regional gene expression.," *Genes, chromosomes & cancer*, vol. 46, no. 10, pp. 936–949, 2007.

[5] E. Michels, J. Vandesompele, K. De Preter, J. Hoebeeck, J. Vermeulen, A. Schramm, J. J. Molenaar, B. Menten, B. Marques, R. L. Stallings, V. Combaret, C. Devalck, A. De Paepe, R. Versteeg, A. Eggert, G. Laureys, N. Van Roy,

and F. Speleman, "ArrayCGH-based classification of neuroblastoma into genomic subgroups.," *Genes, chromosomes & cancer*, vol. 46, no. 12, pp. 1098–1108, 2007.

[6] H. J. Nickerson, K. K. Matthay, R. C. Seeger, G. M. Brodeur, H. Shimada, C. Perez, J. B. Atkinson, M. Selch, R. B. Gerbing, D. O. Stram, and J. Lukens, "Favorable biology and outcome of stage IV-S neuroblastoma with supportive care or minimal therapy: a Children's Cancer Group study," *J. Clin. Oncol.*, vol. 18, pp. 477–486, Feb 2000.

[7] S. L. Cohn, A. D. Pearson, W. B. London, T. Monclair, P. F. Ambros, G. M. Brodeur, A. Faldum, B. Hero, T. Iehara, D. Machin, V. Mosseri, T. Simon, A. Garaventa, V. Castel, K. K. Matthay, and I. T. Force, "The international neuroblastoma risk group (INRG) classification system: an INRG task force report.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 27, no. 2, pp. 289–297, 2009.

[8] A. L. Yu, A. L. Gilman, M. Ozkaynak, W. B. London, S. G. Kreissman, H. X. Chen, M. Smith, B. Anderson, J. G. Villablanca, K. K. Matthay, H. Shimada, S. A. Grupp, R. Seeger, C. Reynolds, A. Buxton, R. A. Reisfeld, S. D. Gillies, S. L. Cohn, J. M. Maris, P. M. Sondel, and C. O. Group, "Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma.," *The New England journal of medicine*, vol. 363, no. 14, pp. 1324–1334, 2010.

[9] T. Perwein, H. Lackner, P. Sovinz, M. Benesch, S. Schmidt, W. Schwinger, and C. Urban, "Survival and late effects in children with stage 4 neuroblastoma.," *Pediatric blood & cancer*, vol. 57, no. 4, pp. 629–635, 2011.

[10] C. Sawyers, "Targeted cancer therapy.," *Nature*, vol. 432, no. 7015, pp. 294–297, 2004.

[11] D. Slamon, L. B, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, T. Fleming, W. Eiermann, J. Wolter, M. Pegram, J. Baselga, and L. Norton, "Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2.," *The New England journal of medicine*, vol. 344, no. 11, pp. 783–792, 2001.

[12] B. Druker, M. Talpaz, D. Resta, B. Peng, E. Buchdunger, J. Ford, N. Lydon, H. Kantarjian, R. Capdeville, O. S, and C. Sawyers, "Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia.," *The New England journal of medicine*, vol. 344, no. 14, pp. 1031–1037, 2001.

[13] S. A. Grupp, M. Kalos, D. Barrett, R. Aplenc, D. L. Porter, S. R. Rheingold, D. T. Teachey, A. Chew, B. Hauck, J. Wright, M. C. Milone, B. L. Levine, and C. H. June, "Chimeric antigen receptor-modified t cells for acute lymphoid leukemia.," *The New England journal of medicine*, vol. 368, no. 16, pp. 1509–1518, 2013.

[14] S. L. Topalian, C. G. Drake, and D. M. Pardoll, "Immune checkpoint blockade: a common denominator approach to cancer therapy.," *Cancer cell*, vol. 27, no. 4, pp. 450–461, 2015.

[15] P. Sharma and J. P. Allison, "Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential.," *Cell*, vol. 161, no. 2, pp. 205–214, 2015.

[16] R. Seeger, G. Brodeur, H. Sather, A. Dalton, S. Siegel, K. Wong, and D. Hammond, "Association of multiple copies of the n-myc oncogene with rapid progression of neuroblastomas.," *The New England journal of medicine*, vol. 313, no. 18, pp. 1111–1116, 1985.

[17] J. Maris, P. White, C. Beltinger, E. Sulman, R. Castleberry, J. Shuster, A. Look, and G. Brodeur, "Significance of chromosome 1p loss of heterozygosity in neuroblastoma.," *Cancer research*, vol. 55, no. 20, pp. 4664–4669, 1995.

[18] H. Caron, P. van Sluis, J. de Kraker, J. Bökkerink, M. Egeler, G. Laureys, R. Slater, A. Westerveld, P. Voûte, and R. Versteeg, "Allelic loss of chromosome 1p as a predictor of unfavorable outcome in patients with neuroblastoma.," *The New England journal of medicine*, vol. 334, no. 4, pp. 225–230, 1996.

[19] N. Bown, S. Cotterill, M. Lastowska, O. S, A. Pearson, D. Plantaz, M. Meddeb, G. Danglot, C. Brinkschmidt, H. Christiansen, G. Laureys, F. Speleman, J. Nicholson, A. Bernheim, D. Betts, J. Vandesompele, and N. Van Roy, "Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma.," *The New England journal of medicine*, vol. 340, no. 25, pp. 1954–1961, 1999.

[20] E. F. Attiyeh, W. B. London, Y. P. Mossé, Q. Wang, C. Winter, D. Khazi, M. P. W, R. C. Seeger, A. Look, H. Shimada, G. M. Brodeur, S. L. Cohn, K. K. Matthay, J. M. Maris, and C. O. Group, "Chromosome 1p and 11q deletions and outcome in neuroblastoma.," *The New England journal of medicine*, vol. 353, no. 21, pp. 2243–2253, 2005.

[21] M. Schwab, K. Alitalo, K. H. Klempnauer, H. E. Varmus, J. M. Bishop, F. Gilbert, G. Brodeur, M. Goldstein, and J. Trent, "Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour," *Nature*, vol. 305, no. 5931, pp. 245–248, 1983.

[22] G. M. Brodeur, R. C. Seeger, M. Schwab, H. E. Varmus, and J. M. Bishop, "Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage," *Science*, vol. 224, pp. 1121–1124, Jun 1984.

[23] P. A. Muller and K. H. Vousden, "Mutant p53 in cancer: new functions and therapeutic opportunities.," *Cancer cell*, vol. 25, no. 3, pp. 304–317, 2014.

[24] J. Imamura, C. R. Bartram, F. Berthold, D. Harms, H. Nakamura, and H. P. Koeffler, "Mutation of the p53 gene in neuroblastoma and its relationship with N-myc amplification," *Cancer Res.*, vol. 53, pp. 4053–4058, Sep 1993.

[25] K. Vogan, M. Bernstein, J. M. Leclerc, L. Brisson, J. Brossard, G. M. Brodeur, J. Pelletier, and P. Gros, "Absence of p53 gene mutations in primary neuroblastomas," *Cancer Res.*, vol. 53, pp. 5269–5273, Nov 1993.

[26] K. Forrester, C. Almoguera, K. Han, W. E. Grizzle, and M. Perucho, "Detection of high incidence of K-ras oncogenes during human colon tumorigenesis," *Nature*, vol. 327, no. 6120, pp. 298–303, 1987.

[27] C. Almoguera, D. Shibata, K. Forrester, J. Martin, N. Arnheim, and M. Perucho, "Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes," *Cell*, vol. 53, pp. 549–554, May 1988.

[28] J. F. Moley, M. B. Brother, S. A. Wells, B. A. Spengler, J. L. Biedler, and G. M. Brodeur, "Low frequency of ras gene mutations in neuroblastomas, pheochromocytomas, and medullary thyroid cancers," *Cancer Res.*, vol. 51, pp. 1596–1599, Mar 1991.

[29] A. Iolascon, M. Badiali, A. Pession, G. Basso, L. Losi, E. Delgiudice, S. Perrotta, S. Cutillo, and G. Tonini, "Rare frequencey of point mutations for codon 12, 13 and 61 of ras gene in italian neuroblastoma," *Int. J. Oncol.*, vol. 3, pp. 529–533, Sep 1993.

[30] D. Trochet, F. Bourdeaut, J. Isabelle, A. Deville, L. de Pontual, G. Schleiermacher,

C. Coze, N. Philip, T. Frébourg, A. Munnich, S. Lyonnet, O. Delattre, and J. Amiel, "Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma.," *American journal of human genetics*, vol. 74, no. 4, pp. 761–764, 2004.

[31] Y. P. Mosse, M. Laudenslager, D. Khazi, A. J. Carlisle, C. L. Winter, E. Rappaport, and J. M. Maris, "Germline PHOX2B mutation in hereditary neuroblastoma.," *American journal of human genetics*, vol. 75, no. 4, pp. 727–730, 2004.

[32] J. Amiel, B. Laudier, A. Tania, H. Trang, L. de Pontual, B. Gener, D. Trochet, H. Etchevers, P. Ray, M. Simonneau, M. Vekemans, A. Munnich, C. Gaultier, and S. Lyonnet, "Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome.," *Nature genetics*, vol. 33, no. 4, pp. 459–461, 2003.

[33] E. Lander, L. Linton, B. Birren, C. Nusbaum, M. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, F. W, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, L. R, M. P, M. K, J. Meldrim, J. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, S. Y, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, M. A, L. Matthews, S. Mercer, S. Milne, J. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. Waterston, R. Wilson, L. Hillier, M. JD, M. Marra, E. Mardis, L. Fulton, A. Chinwalla, K. Pepin, W. Gish, S. Chissoe, M. Wendl, K. Delehaunty, T. Miner, A. Delehaunty, J. Kramer,

L. Cook, R. Fulton, D. Johnson, P. Minx, S. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, and M. Frazier, "Initial sequencing and analysis of the human genome.," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[34] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann,

B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu, "The sequence of the human genome," *Science*, vol. 291, pp. 1304–1351, Feb 2001.

[35] M. L. Metzker, "Sequencing technologies - the next generation.," *Nature reviews. Genetics*, vol. 11, no. 1, pp. 31–46, 2010.

[36] Y. P. P. Mossé, M. Laudenslager, L. Longo, K. A. Cole, A. Wood, E. F. Attiyeh, M. J. Laquaglia, R. Sennett, J. E. Lynch, P. Perri, G. Laureys, F. Speleman,

C. Kim, C. Hou, H. Hakonarson, A. Torkamani, N. J. Schork, G. M. Brodeur, G. P. Tonini, E. Rappaport, M. Devoto, and J. M. Maris, "Identification of ALK as a major familial neuroblastoma predisposition gene.," *Nature*, vol. 455, no. 7215, pp. 930–935, 2008.

[37] J. Isabelle, D. Lequin, L. Brugières, A. Ribeiro, L. de Pontual, V. Combaret, V. Raynal, A. Puisieux, G. Schleiermacher, G. Pierron, V. Dominique, T. Frebourg, J. Michon, S. Lyonnet, J. Amiel, and O. Delattre, "Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma.," *Nature*, vol. 455, no. 7215, pp. 967–970, 2008.

[38] Y. Chen, J. Takita, Y. L. Choi, M. Kato, M. Ohira, M. Sanada, L. Wang, M. Soda, A. Kikuchi, T. Igarashi, A. Nakagawara, Y. Hayashi, H. Mano, and S. Ogawa, "Oncogenic mutations of ALK kinase in neuroblastoma.," *Nature*, vol. 455, no. 7215, pp. 971–974, 2008.

[39] R. E. George, T. Sanda, M. Hanna, S. Fröhling, W. Luther, J. Zhang, Y. Ahn, W. Zhou, W. B. London, M. Patrick, L. Xue, S. Zozulya, V. E. Gregor, T. R. Webb, N. S. Gray, D. Gilliland, L. Diller, H. Greulich, S. W. Morris, M. Meyerson, and A. Look, "Activating mutations in ALK provide a therapeutic target in neuroblastoma.," *Nature*, vol. 455, no. 7215, pp. 975–978, 2008.

[40] M. Soda, Y. L. Choi, M. Enomoto, S. Takada, Y. Yamashita, S. Ishikawa, S.-i. Fujiwara, H. Watanabe, K. Kurashina, H. Hatanaka, M. Bando, S. Ohno, Y. Ishikawa, H. Aburatani, T. Niki, Y. Sohara, Y. Sugiyama, and H. Mano, "Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer.," *Nature*, vol. 448, no. 7153, pp. 561–566, 2007.

[41] E. L. Kwak, Y. J. Bang, D. Camidge, A. T. Shaw, B. Solomon, R. G. Maki,

S. I. H. Ou, B. J. Dezube, P. A. Jänne, D. B. Costa, V. Marileila, W. H. Kim, T. J. Lynch, P. Fidias, H. Stubbs, J. A. Engelman, L. V. Sequist, W. Tan, L. Gandhi, M. Mari, G. C. Wei, S. Shreeve, M. J. Ratain, J. Settleman, J. G. Christensen, D. A. Haber, K. Wilner, R. Salgia, G. I. Shapiro, J. W. Clark, and A. Iafrate, "Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer.," *The New England journal of medicine*, vol. 363, no. 18, pp. 1693–1703, 2010.

[42] Y. P. Mossé, A. Wood, and J. M. Maris, "Inhibition of ALK signaling for cancer therapy.," *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 15, no. 18, pp. 5609–5614, 2009.

[43] Y. P. Mossé, M. S. Lim, S. D. Voss, K. Wilner, K. Ruffner, J. Laliberte, D. Rolland, F. M. Balis, J. M. Maris, B. J. Weigel, A. M. Ingle, C. Ahern, P. C. Adamson, and S. M. Blaney, "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.," *The lancet oncology*, vol. 14, no. 6, pp. 472–480, 2013.

[44] J. J. Molenaar, J. Koster, D. A. Zwijnenburg, P. van Sluis, L. J. Valentijn, I. van der Ploeg, M. Hamdi, J. van Nes, B. A. Westerman, J. van Arkel, M. E. Ebus, F. Haneveld, A. Lakeman, L. Schild, P. Molenaar, P. Stroeken, M. M. van Noesel, I. Ora, E. E. Santo, H. N. Caron, E. M. Westerhout, and R. Versteeg, "Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes.," *Nature*, vol. 483, no. 7391, pp. 589–593, 2012.

[45] N. V. K. Cheung, J. Zhang, C. Lu, M. Parker, A. Bahrami, S. K. Tickoo, A. Heguy, A. S. Pappo, S. Federico, J. Dalton, I. Y. Cheung, L. Ding, R. Fulton, J. Wang, X. Chen, J. Becksfort, J. Wu, C. A. Billups, D. Ellison, E. R. Mardis,

R. K. Wilson, J. R. Downing, M. A. Dyer, and S. J. C. R. H. U. P. C. G. Project, "Association of age at diagnosis and genetic mutations in patients with neuroblastoma.," *JAMA*, vol. 307, no. 10, pp. 1062–1071, 2012.

[46] T. J. Pugh, O. Morozova, E. F. Attiyeh, S. Asgharzadeh, J. S. Wei, D. Auclair, S. L. Carter, K. Cibulskis, M. Hanna, A. Kiezun, J. Kim, M. S. Lawrence, L. Lichenstein, M. Aaron, C. S. Pedamallu, A. H. Ramos, E. Shefler, A. Sivachenko, C. Sougnez, C. Stewart, A. Ally, I. Birol, R. Chiu, R. D. Corbett, M. Hirst, S. D. Jackman, B. Kamoh, A. H. Khodabakshi, M. Krzywinski, A. Lo, R. A. Moore, K. L. Mungall, J. Qian, A. Tam, N. Thiessen, Y. Zhao, K. A. Cole, M. Diamond, S. J. Diskin, Y. P. Mosse, A. C. Wood, L. Ji, R. Sposto, T. Badgett, W. B. London, Y. Moyer, G. J. M, M. A. Smith, J. M. Guidry Auvil, D. S. Gerhard, M. D. Hogarty, S. J. Jones, E. S. Lander, S. B. Gabriel, G. Getz, R. C. Seeger, J. Khan, M. A. Marra, M. Meyerson, and J. M. Maris, "The genetic landscape of high-risk neuroblastoma.," *Nature genetics*, vol. 45, no. 3, pp. 279–284, 2013.

[47] M. Sausen, R. J. Leary, S. Jones, J. Wu, C. Reynolds, X. Liu, A. Blackford, G. Parmigiani, L. A. Diaz, N. Papadopoulos, B. Vogelstein, K. W. Kinzler, V. E. Velculescu, and M. D. Hogarty, "Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma.," *Nature genetics*, vol. 45, no. 1, pp. 12–17, 2013.

[48] M. Capasso, M. Devoto, C. Hou, S. Asgharzadeh, J. T. Glessner, E. F. Attiyeh, Y. P. Mosse, C. Kim, S. J. Diskin, K. A. Cole, K. Bosse, M. Diamond, M. Laudenslager, C. Winter, J. P. Bradfield, R. H. Scott, J. Jagannathan, M. Garris, M. Carmel, W. B. London, R. C. Seeger, S. F. Grant, H. Li, N. Rahman, E. Rappaport, H. Hakonarson, and J. M. Maris, "Common variations in BARD1

influence susceptibility to high-risk neuroblastoma.," *Nature genetics*, vol. 41, no. 6, pp. 718–723, 2009.

[49] J. M. Maris, Y. P. Mosse, J. P. Bradfield, C. Hou, S. Monni, R. H. Scott, S. Asgharzadeh, E. F. Attiyeh, S. J. Diskin, M. Laudenslager, C. Winter, K. A. Cole, J. T. Glessner, C. Kim, E. C. Frackelton, T. Casalunovo, A. W. Eckert, M. Capasso, E. F. Rappaport, M. Carmel, W. B. London, R. C. Seeger, N. Rahman, M. Devoto, S. F. Grant, H. Li, and H. Hakonarson, "Chromosome 6p22 locus associated with clinically aggressive neuroblastoma.," *The New England journal of medicine*, vol. 358, no. 24, pp. 2585–2593, 2008.

[50] K. Wang, S. J. Diskin, H. Zhang, E. F. Attiyeh, C. Winter, C. Hou, R. W. Schnepp, M. Diamond, K. Bosse, P. A. Mayes, J. Glessner, C. Kim, E. Frackelton, M. Garris, Q. Wang, W. Glaberson, R. Chiavacci, L. Nguyen, J. Jagannathan, N. Saeki, H. Sasaki, S. F. Grant, A. Iolascon, Y. P. Mosse, K. A. Cole, H. Li, M. Devoto, M. P. W, W. B. London, M. Capasso, N. Rahman, H. Hakonarson, and J. M. Maris, "Integrative genomics identifies LMO1 as a neuroblastoma oncogene.," *Nature*, vol. 469, no. 7329, pp. 216–220, 2011.

[51] S. J. Diskin, M. Capasso, R. W. Schnepp, K. A. Cole, E. F. Attiyeh, C. Hou, M. Diamond, E. L. Carpenter, C. Winter, H. Lee, J. Jagannathan, V. Latorre, A. Iolascon, H. Hakonarson, M. Devoto, and J. M. Maris, "Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma.," *Nature genetics*, vol. 44, no. 10, pp. 1126–1130, 2012.

[52] S. J. Diskin, M. Capasso, M. Diamond, D. A. Oldridge, K. Conkrite, K. R. Bosse, M. R. Russell, A. Iolascon, H. Hakonarson, M. Devoto, and J. M. Maris, "Rare

variants in TP53 and susceptibility to neuroblastoma.," *Journal of the National Cancer Institute*, vol. 106, no. 4, p. dju047, 2014.

[53] L. B. e. B. Nguyen, S. J. Diskin, M. Capasso, K. Wang, M. A. Diamond, J. Glessner, C. Kim, E. F. Attiyeh, Y. P. Mosse, K. Cole, A. Iolascon, M. Devoto, H. Hakonarson, H. K. Li, and J. M. Maris, "Phenotype restricted genome-wide association study using a gene-centric approach identifies three low-risk neuroblastoma susceptibility loci.," *PLoS genetics*, vol. 7, no. 3, p. e1002026, 2011.

[54] S. J. Diskin, C. Hou, J. T. Glessner, E. F. Attiyeh, M. Laudenslager, K. Bosse, K. Cole, Y. P. P. Mossé, A. Wood, J. E. Lynch, K. Pecor, M. Diamond, C. Winter, K. Wang, C. Kim, E. A. Geiger, M. P. W, A. I. Blakemore, W. B. London, T. H. Shaikh, J. Bradfield, S. F. Grant, H. Li, M. Devoto, E. R. Rappaport, H. Hakonarson, and J. M. Maris, "Copy number variation at 1q21.1 associated with neuroblastoma.," *Nature*, vol. 459, no. 7249, pp. 987–991, 2009.

[55] T. C. Voss and G. L. Hager, "Dynamic regulation of transcriptional states by chromatin and transcription factors.," *Nature reviews. Genetics*, vol. 15, no. 2, pp. 69–81, 2014.

[56] T. Chen and S. Y. Dent, "Chromatin modifiers and remodellers: regulators of cellular differentiation.," *Nature reviews. Genetics*, vol. 15, no. 2, pp. 93–9106, 2014.

[57] C. Plass, S. M. Pfister, A. M. Lindroth, O. Bogatyrova, R. Claus, and P. Lichter, "Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer.," *Nature reviews. Genetics*, vol. 14, no. 11, pp. 765–780, 2013.

[58] L. Erez, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker, "Comprehensive mapping of long-range interactions reveals folding principles of the human genome.," *Science (New York, N.Y.)*, vol. 326, no. 5950, pp. 289–293, 2009.

[59] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden, "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.," *Cell*, vol. 159, no. 7, pp. 1665–1680, 2014.

[60] F. W. Huang, E. Hodis, M. J. Xu, G. V. Kryukov, L. Chin, and L. A. Garraway, "Highly recurrent TERT promoter mutations in human melanoma.," *Science (New York, N.Y.)*, vol. 339, no. 6122, pp. 957–959, 2013.

[61] S. Horn, A. Figl, P. Rachakonda, C. Fischer, A. Sucker, A. Gast, S. Kadel, I. Moll, E. Nagore, K. Hemminki, D. Schadendorf, and R. Kumar, "TERT promoter mutations in familial and sporadic melanoma.," *Science (New York, N.Y.)*, vol. 339, no. 6122, pp. 959–961, 2013.

[62] E. P. Consortium, B. E. Bernstein, E. Birney, I. Dunham, E. D. Green, C. Gunter, and M. Snyder, "An integrated encyclopedia of DNA elements in the human genome.," *Nature*, vol. 489, no. 7414, pp. 57–74, 2012.

[63] B. E. Bernstein, J. A. Stamatoyannopoulos, J. F. Costello, B. Ren, A. Milosavljevic, A. Meissner, M. Kellis, M. A. Marra, A. L. Beaudet, J. R. Ecker, P. J. Farnham, M. Hirst, E. S. Lander, T. S. Mikkelsen, and J. A. Thomson, "The

NIH roadmap epigenomics mapping consortium.," *Nature biotechnology*, vol. 28, no. 10, pp. 1045–1048, 2010.

[64] S. L. Edwards, J. Beesley, J. D. French, and A. M. Dunning, "Beyond GWASs: illuminating the dark road from association to function.," *American journal of human genetics*, vol. 93, no. 5, pp. 779–797, 2013.

[65] M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Sandra, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, and J. A. Stamatoyannopoulos, "Systematic localization of common disease-associated variation in regulatory DNA.," *Science (New York, N.Y.)*, vol. 337, no. 6099, pp. 1190–1195, 2012.

[66] Q. Li, J. H. Seo, B. Stranger, M. Aaron, I. Pe'er, T. Laframboise, M. Brown, S. Tyekucheva, and M. L. Freedman, "Integrative eQTL-based analyses reveal the biology of breast cancer risk loci.," *Cell*, vol. 152, no. 3, pp. 633–641, 2013.

[67] M. M. Pomerantz, N. Ahmadiyeh, L. Jia, P. Herman, M. P. Verzi, H. Doddapaneni, C. A. Beckwith, J. A. Chan, A. Hills, M. Davis, K. Yao, S. M. Kehoe, H. J. Lenz, C. A. Haiman, C. Yan, B. E. Henderson, B. Frenkel, J. Barretina, A. Bass, J. Tabernero, J. Baselga, M. M. Regan, J. Manak, R. Shivdasani, G. A. Coetzee, and M. L. Freedman, "The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer.," *Nature genetics*, vol. 41, no. 8, pp. 882–884, 2009.

[68] N. Ahmadiyeh, M. M. Pomerantz, C. Grisanzio, P. Herman, L. Jia, V. Almendro, H. H. He, M. Brown, X. Liu, M. Davis, J. L. Caswell, C. A. Beckwith, A. Hills,

L. Macconaill, G. A. Coetzee, M. M. Regan, and M. L. Freedman, "8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 21, pp. 9742–9746, 2010.

[69] D. J. Hazelett, S. K. Rhie, M. Gaddis, C. Yan, D. L. Lakeland, S. G. Coetzee, E. consortium, P. consortium, B. E. Henderson, H. Noushmehr, W. Cozen, K. Zsofia, R. A. Eeles, D. F. Easton, C. A. Haiman, W. Lu, P. J. Farnham, and G. A. Coetzee, "Comprehensive functional annotation of 77 prostate cancer risk loci.," *PLoS genetics*, vol. 10, no. 1, p. e1004102, 2014.

[70] K. Lawrenson, Q. Li, S. Kar, J. H. Seo, J. Tyrer, T. J. Spindler, J. Lee, Y. Chen, A. Karst, R. Drapkin, K. K. Aben, A. Hoda, N. Antonenkova, A. O. C. S. Group, H. Baker, E. V. Bandera, Y. Bean, M. W. Beckmann, A. Berchuck, M. Bisogna, L. Bjorge, N. Bogdanova, L. A. Brinton, B. Angela, F. Bruinsma, R. Butzow, I. G. Campbell, K. Carty, C. Jenny, C. Georgia, A. Chen, Z. Chen, L. S. Cook, D. W. Cramer, J. M. Cunningham, C. Cybulski, D. Agnieszka, J. Dennis, E. Dicks, J. A. Doherty, T. Dörk, A. du Bois, M. Dürst, D. Eccles, D. T. Easton, R. P. Edwards, U. Eilber, A. B. Ekici, P. A. Fasching, B. L. Fridley, Y. T. Gao, G. Aleksandra, G. G. Giles, R. Glasspool, E. L. Goode, M. T. Goodman, J. Grownwald, P. Harrington, P. Harter, H. N. Hasmad, A. Hein, F. Heitz, M. A. Hildebrandt, P. Hillemanns, E. Hogdall, C. Hogdall, S. Hosono, E. S. Iversen, A. Jakubowska, P. James, A. Jensen, B. T. Ji, B. Y. Karlan, S. Kruger Kjaer, L. E. Kelemen, M. Kellar, J. L. Kelley, L. A. Kiemeney, C. Krakstad, J. Kupryjanczyk, D. Lambrechts, S. Lambrechts, N. D. Le, A. W. Lee, S. Lele, A. Leminen, J. Lester, D. A. Levine, D. Liang, J. Lissowska, K. Lu, J. Lubinski, L. Lundvall, L. F. Massuger, K. Matsuo, M. Valerie, M. J. R, H. Nevanlinna, M. Ian, and U. Menon,

"Cis-eQTL analysis and functional validation of candidate susceptibility genes for high-grade serous ovarian cancer.," *Nature communications*, vol. 6, p. 8234, 2015.

[71] M. L. Freedman, A. N. Monteiro, S. A. Gayther, G. A. Coetzee, A. Risch, C. Plass, G. Casey, M. De Biasi, C. Carlson, D. Duggan, M. James, P. Liu, J. W. Tichelaar, H. G. Vikis, M. You, and I. G. Mills, "Principles for the post-GWAS functional characterization of cancer risk loci.," *Nature genetics*, vol. 43, no. 6, pp. 513–518, 2011.

[72] D. A. Oldridge, A. C. Wood, W. Nina, I. Crimmins, R. Sussman, C. Winter, M. L. D, M. Diamond, L. S. Hart, S. Zhu, A. D. Durbin, B. J. Abraham, L. Anders, L. Tian, S. Zhang, J. S. Wei, J. Khan, K. Bramlett, N. Rahman, M. Capasso, A. Iolascon, D. S. Gerhard, J. M. Guidry Auvil, R. A. Young, H. Hakonarson, S. J. Diskin, A. Thomas Look, and J. M. Maris, "Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism.," *Nature*, 2015.

[73] J. M. Matthews, K. Lester, S. Joseph, and D. J. Curtis, "LIM-domain-only proteins in cancer.," *Nature reviews. Cancer*, vol. 13, no. 2, pp. 111–122, 2013.

[74] A. Mathelier, X. Zhao, A. W. Zhang, F. Parcy, W. Rebecca, D. J. Arenillas, S. Buchman, C.-y. Y. Chen, A. Chou, H. Ienasescu, J. Lim, C. Shyr, G. Tan, M. Zhou, B. Lenhard, A. Sandelin, and W. W. Wasserman, "JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles.," *Nucleic acids research*, vol. 42, no. Database issue, pp. D142–D147, 2014.

[75] Y. Chen, D. L. Bates, R. Dey, P. H. Chen, A. C. Machado, L. I. A, R. Rohs,

and L. Chen, "DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation.," *Cell reports*, vol. 2, no. 5, pp. 1197–1206, 2012.

[76] A. Zahirieh, M. Nesbit, A. Ali, K. Wang, N. He, M. Stangou, G. Bamichas, K. Sombolos, R. V. Thakker, and Y. Pei, "Functional analysis of a novel GATA3 mutation in a family with the hypoparathyroidism, deafness, and renal dysplasia syndrome.," *The Journal of clinical endocrinology and metabolism*, vol. 90, no. 4, pp. 2445–2450, 2005.

[77] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne, "The protein data bank.," *Nucleic acids research*, vol. 28, no. 1, pp. 235–242, 2000.

[78] W. Deng, J. Lee, H. Wang, J. Miller, A. Reik, P. D. Gregory, A. Dean, and G. A. Blobel, "Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor.," *Cell*, vol. 149, no. 6, pp. 1233–1244, 2012.

[79] M. P. Creyghton, A. W. Cheng, G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch, "Histone H3K27ac separates active from poised enhancers and predicts developmental state.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 50, pp. 21931–21936, 2010.

[80] D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, S. Violaine, A. A. Sigova, H. A. Hoke, and R. A. Young, "Super-enhancers in the control of cell identity and disease.," *Cell*, vol. 155, no. 4, pp. 934–947, 2013.

[81] T. Sanda, L. N. Lawton, M. Barrasa, Z. P. Fan, H. Kohlhammer, A. Gutierrez, W. Ma, J. Tatarek, Y. Ahn, M. A. Kelliher, C. H. Jamieson, L. M. Staudt, R. A. Young, and A. Look, "Core transcriptional regulatory circuit controlled by the TAL1 complex in human t cell acute lymphoblastic leukemia.," *Cancer cell*, vol. 22, no. 2, pp. 209–221, 2012.

[82] E. Chipumuro, E. Marco, C. L. Christensen, N. Kwiatkowski, T. Zhang, C. M. Hatheway, B. J. Abraham, B. Sharma, C. Yeung, A. Altabef, P. Antonio, K. K. Wong, G. C. Yuan, N. S. Gray, R. A. Young, and R. E. George, "CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer.," *Cell*, vol. 159, no. 5, pp. 1126–1139, 2014.

[83] S. C. Parker, M. L. Stitzel, D. Taylor, J. M. Orozco, M. R. Erdos, J. A. Akiyama, K. L. van Bueren, P. S. Chines, N. Narisu, N. C. S. Program, B. L. Black, A. Visel, L. A. Pennacchio, F. S. Collins, N. I. of Health Intramural Sequencing Center Comparative Sequencing Program Authors, and N. C. S. P. Authors, "Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 44, pp. 17921–17926, 2013.

[84] A. Puissant, S. M. Frumm, G. Alexe, C. F. Bassil, J. Qi, Y. H. Chanthery, E. A. Nekritz, R. Zeid, W. C. Gustafson, P. Greninger, M. J. Garnett, M. Ultan, C. H. Benes, A. L. Kung, W. A. Weiss, J. E. Bradner, and K. Stegmaier, "Targeting MYCN in neuroblastoma by BET bromodomain inhibition.," *Cancer discovery*, vol. 3, no. 3, pp. 308–323, 2013.

[85] I. Sur, S. Tuupanen, T. Whitington, L. A. Aaltonen, and J. Taipale, "Lessons

from functional analysis of genome-wide association studies.," *Cancer research*, vol. 73, no. 14, pp. 4180–4184, 2013.

[86] W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, and R. A. Young, "Master transcription factors and mediator establish super-enhancers at key cell identity genes.," *Cell*, vol. 153, no. 2, pp. 307–319, 2013.

[87] T. O. Henderson, S. Bhatia, N. Pinto, W. B. London, M. Patrick, C. Crotty, C. L. Sun, and S. L. Cohn, "Racial and ethnic disparities in risk and survival in children with neuroblastoma: a children's oncology group study.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 29, no. 1, pp. 76–82, 2011.

[88] A. Monteiro and M. Freedman, "Lessons from postgenome-wide association studies: functional analysis of cancer predisposition loci.," *Journal of internal medicine*, vol. 274, no. 5, pp. 414–424, 2013.

[89] R. A. Stewart, J. S. Lee, M. Lachnit, A. Look, J. P. Kanki, and P. D. Henion, "Studying peripheral sympathetic nervous system development and neuroblastoma in zebrafish.," *Methods in cell biology*, vol. 100, pp. 127–152, 2010.

[90] S. Zhu, J. S. Lee, F. Guo, J. Shin, P. A. R, J. L. Kutok, S. J. Rodig, D. S. Neuberg, D. Helman, H. Feng, R. A. Stewart, W. Wang, R. E. George, J. P. Kanki, and A. Look, "Activated ALK collaborates with MYCN in neuroblastoma pathogenesis.," *Cancer cell*, vol. 21, no. 3, pp. 362–373, 2012.

[91] J. Molenaar, M. Ebus, J. Koster, E. Santo, D. Geerts, R. Versteeg, and H. Caron, "Cyclin d1 is a direct transcriptional target of GATA3 in neuroblastoma tumor cells.," *Oncogene*, vol. 29, no. 18, pp. 2739–2745, 2010.

[92] H. Shi and H.-F. Ding, "Gata3 regulation of human neuroblastoma stem cell activity," *The FASEB Journal*, vol. 23, no. 1_MeetingAbstracts, pp. 740–14, 2009.

[93] H. Peng, X. X. Ke, R. Hu, L. Yang, H. Cui, and Y. Wei, "Essential role of GATA3 in regulation of differentiation and cell proliferation in SK-N-SH neuroblastoma cells.," *Molecular medicine reports*, vol. 11, no. 2, pp. 881–886, 2015.

[94] N. K. Wilson, S. D. Foster, X. Wang, K. Knezevic, J. Schütte, P. Kaimakis, P. M. Chilarska, S. Kinston, W. H. Ouwehand, E. Dzierzak, J. E. Pimanda, M. F. de Bruijn, and B. Göttgens, "Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators.," *Cell stem cell*, vol. 7, no. 4, pp. 532–544, 2010.

[95] D. Lee, R. Karchin, and M. A. Beer, "Discriminative prediction of mammalian enhancers from DNA sequence.," *Genome research*, vol. 21, no. 12, pp. 2167–2180, 2011.

[96] M. Ghandi, D. Lee, M. Morteza, and M. A. Beer, "Enhanced regulatory sequence prediction using gapped k-mer features.," *PLoS computational biology*, vol. 10, no. 7, p. e1003711, 2014.

[97] D. Lee, D. U. Gorkin, M. Baker, B. J. Strober, A. L. Asoni, M. A. S, and M. A. Beer, "A method to predict the impact of regulatory variants from DNA sequence.," *Nature genetics*, vol. 47, no. 8, pp. 955–961, 2015.

[98] J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf, "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.," *Nature methods*, vol. 10, no. 12, pp. 1213–1218, 2013.

[99] K. Tano, R. Mizuno, T. Okada, R. Rakwal, J. Shibato, Y. Masuo, K. Ijiri, and N. Akimitsu, "MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes.," *FEBS letters*, vol. 584, no. 22, pp. 4575–4580, 2010.

[100] T. Gutschner, M. Hämmerle, M. Eissmann, J. Hsu, Y. Kim, G. Hung, A. Revenko, G. Arun, M. Stentrup, M. Gross, M. Zörnig, M. AR, D. L. Spector, and S. Diederichs, "The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells.," *Cancer research*, vol. 73, no. 3, pp. 1180–1189, 2013.

[101] B. Hero, T. Simon, R. Spitz, K. Ernestus, A. K. Gnekow, H. G. Scheel-Walter, D. Schwabe, F. H. Schilling, G. Benz-Bohm, and F. Berthold, "Localized infant neuroblastomas often show spontaneous regression: results of the prospective trials NB95-S and NB97," *J. Clin. Oncol.*, vol. 26, pp. 1504–1510, Mar 2008.

[102] T. Simon, F. Berthold, A. Borkhardt, B. Kremens, B. De Carolis, and B. Hero, "Treatment and outcomes of patients with relapsed, high-risk neuroblastoma: results of german trials.," *Pediatric blood & cancer*, vol. 56, no. 4, pp. 578–583, 2011.

[103] J. M. Maris, "Recent advances in neuroblastoma.," *The New England journal of medicine*, vol. 362, no. 23, pp. 2202–2211, 2010.

[104] G. Schleiermacher, N. Javanmardi, V. Bernard, Q. Leroy, J. Cappo, T. Rio Frio, G. Pierron, E. Lapouble, V. Combaret, F. Speleman, B. de Wilde, A. Djos, I. Ora, F. Hedborg, C. Träger, B. M. Holmqvist, J. Abrahamsson, M. Peuchmaur, J. Michon, J. Isabelle, P. Kogner, O. Delattre, and T. Martinsson, "Emergence of new ALK mutations at relapse of neuroblastoma.," *Journal of clinical oncology*

: official journal of the American Society of Clinical Oncology, vol. 32, no. 25, pp. 2727–2734, 2014.

[105] D. Taggart, S. Dubois, and K. K. Matthay, "Radiolabeled metaiodobenzylguanidine for imaging and therapy of neuroblastoma," *Q J Nucl Med Mol Imaging*, vol. 52, pp. 403–418, Dec 2008.

[106] T. F. Eleveld, D. A. Oldridge, V. Bernard, J. Koster, L. C. Daage, S. J. Diskin, L. Schild, N. B. Bentahar, A. Bellini, M. Chicard, E. Lapouble, V. Combaret, L. Patricia, J. Michon, T. J. Pugh, L. S. Hart, J. Rader, E. F. Attiyeh, J. S. Wei, S. Zhang, A. Naranjo, G. J. M, M. D. Hogarty, S. Asgharzadeh, M. A. Smith, J. M. Guidry Auvil, T. B. Watkins, D. A. Zwijnenburg, M. E. Ebus, P. van Sluis, A. Hakkert, E. van Wezel, C. van der Schoot, E. M. Westerhout, J. H. Schulte, G. A. Tytgat, M. Dolman, J. Isabelle, D. S. Gerhard, H. N. Caron, O. Delattre, J. Khan, R. Versteeg, G. Schleiermacher, J. J. Molenaar, and J. M. Maris, "Relapsed neuroblastomas show frequent RAS-MAPK pathway mutations.," *Nature genetics*, vol. 47, no. 8, pp. 864–871, 2015.

[107] S. L. Carter, K. Cibulskis, E. Helman, M. Aaron, H. Shen, T. Zack, P. W. Laird, R. C. Onofrio, W. Winckler, B. A. Weir, R. Beroukhim, D. Pellman, D. A. Levine, E. S. Lander, M. Meyerson, and G. Getz, "Absolute quantification of somatic DNA alterations in human cancer.," *Nature biotechnology*, vol. 30, no. 5, pp. 413–421, 2012.

[108] D. A. Landau, S. L. Carter, P. Stojanov, M. Aaron, K. Stevenson, M. S. Lawrence, C. Sougnez, C. Stewart, A. Sivachenko, L. Wang, Y. Wan, W. Zhang, S. A. Shukla, A. Vartanov, S. M. Fernandes, G. Saksena, K. Cibulskis, B. Tesar, S. Gabriel, N. Hacohen, M. Meyerson, E. S. Lander, D. Neuberg, J. R. Brown, G. Getz, and

C. J. Wu, "Evolution and impact of subclonal mutations in chronic lymphocytic leukemia.," *Cell*, vol. 152, no. 4, pp. 714–726, 2013.

[109] F. Favero, T. Joshi, A. Marquard, N. Birkbak, M. Krzystanek, Q. Li, Z. Szallasi, and A. Eklund, "Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data.," *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, 2014.

[110] S. M. Boca, K. W. Kinzler, V. E. Velculescu, B. Vogelstein, and G. Parmigiani, "Patient-oriented gene set analysis for cancer mutation data.," *Genome biology*, vol. 11, no. 11, p. R112, 2010.

[111] P. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, and M. R. Stratton, "A census of human cancer genes.," *Nature reviews. Cancer*, vol. 4, no. 3, pp. 177–183, 2004.

[112] S. C. Bresler, D. A. Weiser, P. J. Huwe, J. H. Park, K. Krytska, H. Ryles, M. Laudenslager, E. F. Rappaport, A. C. Wood, M. P. W, M. D. Hogarty, W. B. London, R. Radhakrishnan, M. A. Lemmon, and Y. P. P. Mossé, "ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma.," *Cancer cell*, vol. 26, no. 5, pp. 682–694, 2014.

[113] M. Hölzel, S. Huang, J. Koster, I. Ora, A. Lakeman, H. Caron, W. Nijkamp, J. Xie, T. Callens, S. Asgharzadeh, R. C. Seeger, L. Messiaen, R. Versteeg, and R. Bernards, "NF1 is a tumor suppressor in neuroblastoma that determines retinoic acid response and disease outcome.," *Cell*, vol. 142, no. 2, pp. 218–229, 2010.

[114] M. Tartaglia, S. Martinelli, L. Stella, G. Bocchinfuso, E. Flex, V. Cordeddu, G. Zampino, I. v. Burgt, A. Palleschi, T. C. Petrucci, M. Sorcini, C. Schoch,

R. Foa, P. D. Emanuel, and B. D. Gelb, "Diversity and functional consequences of germline and somatic PTPN11 mutations in human disease.," *American journal of human genetics*, vol. 78, no. 2, pp. 279–290, 2006.

[115] C. Jane, O. Kieran, K. M. Wood, C. C. Challen, A. G. Baker, J. R. Board, L. Evans, M. Cole, N. V. K. Cheung, J. Boos, G. Köhler, I. Leuschner, A. D. Pearson, J. Lunec, and D. A. Tweddle, "High frequency of p53/MDM2/p14ARF pathway abnormalities in relapsed neuroblastoma.," *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 16, no. 4, pp. 1108–1118, 2010.

[116] M. J. Garnett, E. J. Edelman, S. J. Heidorn, C. D. Greenman, A. Dastur, K. W. Lau, P. Greninger, I. Thompson, X. Luo, J. Soares, Q. Liu, F. Iorio, D. Surdez, L. Chen, R. J. Milano, G. R. Bignell, A. T. Tam, H. Davies, J. A. Stevenson, S. Barthorpe, S. R. Lutz, F. Kogera, K. Lawrence, M. Anne, X. Mitropoulos, T. Mironenko, H. Thi, L. Richardson, W. Zhou, F. Jewitt, T. Zhang, O. Patrick, J. L. Boisvert, S. Price, W. Hur, W. Yang, X. Deng, A. Butler, H. G. Choi, J. W. Chang, J. Baselga, I. Stamenkovic, J. A. Engelman, S. V. Sharma, O. Delattre, S. Julio, N. S. Gray, J. Settleman, P. Futreal, D. A. Haber, M. R. Stratton, S. Ramaswamy, M. Ultan, and C. H. Benes, "Systematic identification of genomic markers of drug sensitivity in cancer cells.," *Nature*, vol. 483, no. 7391, pp. 570–575, 2012.

[117] G. Schleiermacher, J. Isabelle, A. Ribeiro, J. Klijanienko, J. Couturier, G. Pierron, V. Mosseri, A. Valent, N. Auger, D. Plantaz, H. Rubie, V. Dominique, F. Bourdeaut, V. Combaret, C. Bergeron, J. Michon, and O. Delattre, "Accumulation of segmental alterations determines progression in neuroblastoma.," *Journal of*

clinical oncology : official journal of the American Society of Clinical Oncology, vol. 28, no. 19, pp. 3122–3130, 2010.

[118] G. V. Long, C. Fung, A. M. Menzies, G. M. Pupo, M. S. Carlino, J. Hyman, H. Shahheydari, V. Tembe, J. F. Thompson, R. P. Saw, J. Howle, N. K. Hayward, P. Johansson, R. A. Scolyer, R. F. Kefford, and H. Rizos, "Increased MAPK reactivation in early resistance to dabrafenib/trametinib combination therapy of BRAF-mutant metastatic melanoma.," *Nature communications*, vol. 5, p. 5694, 2014.

[119] T. Berry, W. Luther, N. Bhatnagar, Y. Jamin, E. Poon, T. Sanda, D. Pei, B. Sharma, W. R. Vetharoy, A. Hallsworth, Z. Ahmad, K. Barker, L. Moreau, H. Webber, W. Wang, Q. Liu, P. Antonio, S. Rodig, N. K. Cheung, F. Raynaud, B. Hallberg, S. P. Robinson, N. S. Gray, A. D. Pearson, S. A. Eccles, L. Chesler, and R. E. George, "The ALK(F1174L) mutation potentiates the oncogenic activity of MYCN in neuroblastoma.," *Cancer cell*, vol. 22, no. 1, pp. 117–130, 2012.

[120] N. F. Moore, A. M. Azarova, N. Bhatnagar, K. N. Ross, L. E. Drake, S. Frumm, Q. S. Liu, A. L. Christie, T. Sanda, L. Chesler, A. L. Kung, N. S. Gray, K. Stegmaier, and R. E. George, "Molecular rationale for the use of PI3K/AKT/mTOR pathway inhibitors in combination with crizotinib in ALK-mutated neuroblastoma.," *Oncotarget*, vol. 5, no. 18, pp. 8737–8749, 2014.

[121] A. T. Shaw, D. W. Kim, R. Mehra, D. S. Tan, E. Felip, L. Q. Chow, D. Camidge, J. Vansteenkiste, S. Sharma, T. De Pas, G. J. Riely, B. J. Solomon, J. Wolf, M. Thomas, M. Schuler, G. Liu, A. Santoro, Y. Y. Lau, M. Goldwasser, A. L. Boral, and J. A. Engelman, "Ceritinib in ALK-rearranged non-small-cell lung

cancer.," *The New England journal of medicine*, vol. 370, no. 13, pp. 1189–1197, 2014.

[122] J. Rader, M. R. Russell, L. S. Hart, M. S. Nakazawa, L. T. Belcastro, D. Martinez, Y. Li, E. L. Carpenter, E. F. Attiyeh, S. J. Diskin, S. Kim, S. Parasuraman, G. Caponigro, R. W. Schnepp, A. C. Wood, B. Pawel, K. A. Cole, and J. M. Maris, "Dual CDK4/CDK6 inhibition induces cell-cycle arrest and senescence in neuroblastoma.," *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 19, no. 22, pp. 6173–6182, 2013.

[123] E. Barbieri, P. Mehta, Z. Chen, L. Zhang, A. Slack, S. Berg, and J. M. Shohet, "MDM2 inhibition sensitizes neuroblastoma to chemotherapy-induced apoptotic cell death.," *Molecular cancer therapeutics*, vol. 5, no. 9, pp. 2358–2365, 2006.

[124] T. Van Maerken, F. Speleman, J. Vermeulen, I. Lambertz, S. De Clercq, E. De Smet, N. Yigit, V. Coppens, J. Philippé, A. De Paepe, J. C. Marine, and J. Vandesompele, "Small-molecule MDM2 antagonists as a new therapy concept for neuroblastoma.," *Cancer research*, vol. 66, no. 19, pp. 9646–9655, 2006.

[125] A. Schramm, J. Köster, Y. Assenov, K. Althoff, M. Peifer, E. Mahlow, A. Odersky, D. Beisser, C. Ernst, A. G. Henssen, H. Stephan, C. Schröder, L. Heukamp, A. Engesser, Y. Kahlert, J. Theissen, B. Hero, F. Roels, J. Altmüller, P. Nürnberg, K. Astrahantseff, C. Gloeckner, K. De Preter, C. Plass, S. Lee, H. N. Lode, K. O. Henrich, M. Gartlgruber, F. Speleman, P. Schmezer, F. Westermann, S. Rahmann, M. Fischer, A. Eggert, and J. H. Schulte, "Mutational dynamics between primary and relapse neuroblastomas.," *Nature genetics*, vol. 47, no. 8, pp. 872–877, 2015.

[126] D. D. Shao, W. Xue, E. B. Krall, A. Bhutkar, F. Piccioni, X. Wang, A. C. Schinzel, S. Sood, J. Rosenbluh, J. W. Kim, Y. Zwang, T. M. Roberts, D. E. Root, T. Jacks, and W. C. Hahn, "KRAS and YAP1 converge to regulate EMT and tumor survival.," *Cell*, vol. 158, no. 1, pp. 171–184, 2014.

[127] A. Kapoor, W. Yao, H. Ying, S. Hua, A. Liewen, Q. Wang, Y. Zhong, C. J. Wu, A. Sadanandam, B. Hu, Q. Chang, G. C. Chu, A. Ramsey, S. Jiang, H. Xia, F. Eliot, C. Lim, G. I. Horwitz, A. Viale, P. Pettazzoni, N. Sanchez, H. Wang, A. Protopopov, J. Zhang, T. Heffernan, R. L. Johnson, L. Chin, Y. Wang, G. Draetta, and D. R. A, "Yap1 activation enables bypass of oncogenic kras addiction in pancreatic cancer.," *Cell*, vol. 158, no. 1, pp. 185–197, 2014.

[128] J. E. Yeh, P. A. Toniolo, and D. A. Frank, "Targeting transcription factors: promising new strategies for cancer therapy," *Curr Opin Oncol*, vol. 25, pp. 652–658, Nov 2013.

[129] N. Krug, J. M. Hohlfeld, A. M. Kirsten, O. Kornmann, K. M. Beeh, D. Kappeler, S. Korn, S. Ignatenko, W. Timmer, C. Rogon, J. Zeitvogel, N. Zhang, J. Bille, U. Homburg, A. Turowska, C. Bachert, T. Werfel, R. Buhl, J. Renz, H. Garn, and H. Renz, "Allergen-induced asthmatic responses modified by a GATA3-specific DNAzyme.," *The New England journal of medicine*, vol. 372, no. 21, pp. 1987–1995, 2015.

[130] J. Chambers and T. H. Rabbitts, "Lmo2 at 25 years: a paradigm of chromosomal translocation proteins," *Open biology*, vol. 5, no. 6, p. 150062, 2015.

[131] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.," *Science (New York, N.Y.)*, vol. 337, no. 6096, pp. 816–821, 2012.

[132] F. Ran, P. D. Hsu, C. Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang, and F. Zhang, "Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity.," *Cell*, vol. 154, no. 6, pp. 1380–1389, 2013.

[133] L. Cong, F. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, "Multiplex genome engineering using CRISPR/Cas systems.," *Science (New York, N.Y.)*, vol. 339, no. 6121, pp. 819–823, 2013.

[134] T. J. Cradick, E. J. Fine, C. J. Antico, and G. Bao, "CRISPR/Cas9 systems targeting -globin and CCR5 genes have substantial off-target activity.," *Nucleic acids research*, vol. 41, no. 20, pp. 9584–9592, 2013.

[135] I. B. Hilton, D. A. M, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy, and C. A. Gersbach, "Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers.," *Nature biotechnology*, vol. 33, no. 5, pp. 510–517, 2015.

[136] N. A. Kearns, H. Pham, B. Tabak, R. M. Genga, N. J. Silverstein, M. Garber, and R. Maehr, "Functional annotation of native enhancers with a cas9-histone demethylase fusion.," *Nature methods*, vol. 12, no. 5, pp. 401–403, 2015.

[137] H. Singh, D. L. Longo, and B. A. Chabner, "Improving prospects for targeting RAS.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 33, no. 31, pp. 3650–3659, 2015.

[138] M. Puyol, A. Martín, P. Dubus, F. Mulero, P. Pizcueta, G. Khan, C. Guerra, D. Santamaría, and M. Barbacid, "A synthetic lethal interaction between K-

Ras oncogenes and cdk4 unveils a therapeutic strategy for non-small cell lung carcinoma.," *Cancer cell*, vol. 18, no. 1, pp. 63–73, 2010.

[139] L. N. Kwong, J. C. Costello, H. Liu, S. Jiang, T. L. Helms, A. E. Langsdorf, D. Jakubosky, G. Genovese, F. L. Muller, J. H. Jeong, R. P. Bender, G. C. Chu, K. T. Flaherty, J. A. Wargo, J. J. Collins, and L. Chin, "Oncogenic NRAS signaling differentially regulates survival and proliferation in melanoma.," *Nature medicine*, vol. 18, no. 10, pp. 1503–1510, 2012.

[140] J. M. Ostrem, U. Peters, M. L. Sos, J. A. Wells, and K. M. Shokat, "K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions.," *Nature*, vol. 503, no. 7477, pp. 548–551, 2013.

[141] H. Thompson, "US national cancer institute's new ras project targets an old foe.," *Nature medicine*, vol. 19, no. 8, pp. 949–950, 2013.

[142] N. E. Navin, "The first five years of single-cell cancer genomics and beyond.," *Genome research*, vol. 25, no. 10, pp. 1499–1507, 2015.

[143] V. Latorre, S. J. Diskin, M. A. Diamond, H. Zhang, H. Hakonarson, J. M. Maris, and M. Devoto, "Replication of neuroblastoma SNP association at the BARD1 locus in African-Americans.," *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, vol. 21, no. 4, pp. 658–663, 2012.

[144] M. R. Mansour, B. J. Abraham, L. Anders, A. Berezovskaya, A. Gutierrez, A. D. Durbin, J. Etchin, L. Lawton, S. E. Sallan, L. B. Silverman, M. L. Loh, S. P. Hunger, T. Sanda, R. A. Young, and A. Look, "Oncogene regulation. an oncogenic super-enhancer formed through somatic mutation of a noncoding

intergenic element.," *Science (New York, N.Y.)*, vol. 346, no. 6215, pp. 1373–1377, 2014.

[145] T. I. Lee, S. E. Johnstone, and R. A. Young, "Chromatin immunoprecipitation and microarray-based analysis of protein location.," *Nature protocols*, vol. 1, no. 2, pp. 729–748, 2006.

[146] A. Marson, S. S. Levine, M. F. Cole, G. M. Frampton, T. Brambrink, S. Johnstone, M. G. Guenther, W. K. Johnston, M. Wernig, J. Newman, J. Calabrese, L. M. Dennis, T. L. Volkert, S. Gupta, J. Love, N. Hannett, P. A. Sharp, D. P. Bartel, R. Jaenisch, and R. A. Young, "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.," *Cell*, vol. 134, no. 3, pp. 521–533, 2008.

[147] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.," *Genome biology*, vol. 10, no. 3, p. R25, 2009.

[148] W. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler, "The human genome browser at UCSC.," *Genome research*, vol. 12, no. 6, pp. 996–991006, 2002.

[149] Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. Liu, "Model-based analysis of ChIP-Seq (MACS).," *Genome biology*, vol. 9, no. 9, p. R137, 2008.

[150] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and . G. P. D. P. Subgroup, "The sequence Alignment/Map format and SAMtools.," *Bioinformatics (Oxford, England)*, vol. 25, no. 16, pp. 2078–2079, 2009.

[151] J. Lovén, H. A. Hoke, C. Y. Lin, A. Lau, D. A. Orlando, C. R. Vakoc, J. E. Bradner, T. I. Lee, and R. A. Young, "Selective inhibition of tumor oncogenes by disruption of super-enhancers.," *Cell*, vol. 153, no. 2, pp. 320–334, 2013.

[152] Y. Nakatani and V. Ogryzko, "Immunoaffinity purification of mammalian protein complexes.," *Methods in enzymology*, vol. 370, pp. 430–444, 2003.

[153] A. D. Durbin, G. R. Somers, M. Forrester, M. Pienkowska, G. E. Hannigan, and D. Malkin, "JNK1 determines the oncogenic or tumor-suppressive activity of the integrin-linked kinase in human rhabdomyosarcoma.," *The Journal of clinical investigation*, vol. 119, no. 6, pp. 1558–1570, 2009.

[154] R. Drmanac, A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns, B. G. Kermani, P. Carnevali, I. Nazarenko, G. B. Nilsen, G. Yeung, F. Dahl, A. Fernandez, B. Staker, K. P. Pant, J. Baccash, A. P. Borcherding, A. Brownley, R. Cedeno, L. Chen, D. Chernikoff, A. Cheung, R. Chirita, B. Curson, J. C. Ebert, C. R. Hacker, R. Hartlage, B. Hauser, S. Huang, Y. Jiang, V. Karpinchyk, M. Koenig, C. Kong, T. Landers, C. Le, J. Liu, M. C. E, M. Morenzoni, R. E. Morey, K. Mutch, H. Perazich, K. Perry, B. A. Peters, J. Peterson, C. L. Pethiyagoda, K. Pothuraju, C. Richter, A. M. Rosenbaum, S. Roy, J. Shafto, U. Sharanhovich, K. W. Shannon, C. G. Sheppy, M. Sun, J. V. Thakuria, A. Tran, D. Vu, A. W. Zaranek, X. Wu, S. Drmanac, A. R. Oliphant, W. C. Banyai, B. Martin, D. G. Ballinger, G. M. Church, and C. A. Reid, "Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays.," *Science (New York, N.Y.)*, vol. 327, no. 5961, pp. 78–81, 2010.

[155] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J.

Jones, and M. A. Marra, "Circos: an information aesthetic for comparative genomics.," *Genome research*, vol. 19, no. 9, pp. 1639–1645, 2009.

[156] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform.," *Bioinformatics (Oxford, England)*, vol. 25, no. 14, pp. 1754–1760, 2009.

[157] M. Aaron, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and D. M. A, "The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.," *Genome research*, vol. 20, no. 9, pp. 1297–1303, 2010.

[158] K. Cibulskis, M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E. S. Lander, and G. Getz, "Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.," *Nature biotechnology*, vol. 31, no. 3, pp. 213–219, 2013.

[159] K. Wang, M. Li, and H. Hakonarson, "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.," *Nucleic acids research*, vol. 38, no. 16, p. e164, 2010.

[160] . G. P. Consortium, G. R. Abecasis, A. Auton, L. D. Brooks, D. M. A, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, and M. G. A, "An integrated map of genetic variation from 1,092 human genomes.," *Nature*, vol. 491, no. 7422, pp. 56–65, 2012.

[161] T. Rausch, T. Zichner, A. Schlattl, A. M. Stütz, V. Benes, and J. O. Korbel, "DELLY: structural variant discovery by integrated paired-end and split-read analysis.," *Bioinformatics (Oxford, England)*, vol. 28, no. 18, pp. i333–i339, 2012.

[162] A. Roth, J. Khattra, D. Yap, A. Wan, E. Laks, J. Biele, G. Ha, S. Aparicio, B. Alexandre, and S. P. Shah, "PyClone: statistical inference of clonal population structure in cancer.," *Nature methods*, vol. 11, no. 4, pp. 396–398, 2014.

[163] F. Lamers, L. Schild, J. Koster, R. Versteeg, H. N. Caron, and J. J. Molenaar, "Targeted BIRC5 silencing using YM155 causes cell death in neuroblastoma cells with low ABCB1 expression.," *European journal of cancer (Oxford, England : 1990)*, vol. 48, no. 5, pp. 763–771, 2012.

[164] A. Gnirke, A. Melnikov, J. Maguire, P. Rogov, L. E. M, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander, and C. Nusbaum, "Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing.," *Nature biotechnology*, vol. 27, no. 2, pp. 182–189, 2009.