Dynamic changes of genomic and transcriptomic tumor diversity during melanoma progression

Ariane L. Moore^{1,2}, Christos Dimitrakopoulos^{1,2}, Stefanie Flückiger-Mangual³, Lukas Frischknecht³, Christian Beisel¹, Wilhelm Krek^{3†}, Niko Beerenwinkel^{1,2*}

¹ Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel, Switzerland

² SIB Swiss Institute of Bioinformatics, Mattenstrasse 26, 4058 Basel, Switzerland

³ Institute for Molecular Health Sciences, ETH Zurich, Otto-Stern-Weg 7, 8093 Zurich, Switzerland

† Deceased 29 August 2018

* Correspondence: niko.beerenwinkel@bsse.ethz.ch

Abstract

Metastatic melanoma is an aggressive disease that is notoriously difficult to treat. Here, we used a BRAF^{V600E}-driven mouse model to study melanoma progression and metastasis in the absence of extrinsic mutagenic factors such as UV radiation. The tumors were grown for different time periods in seven different mice such that disease progression was tracked over time. We sequenced the whole exome and whole transcriptome of 16 melanoma samples from these mice. The timed samples revealed that many genes are increasingly upregulated during progression, such as genes involved in extracellular matrix organization, and *Vegfc*, a factor that induces lymphatic vessel growth and metastases to lymph nodes, which was identified as an important signaling hub in our analysis. The *Vegfc* gene interaction network highlighted multiple genes from the MAPK pathway that are increasingly deregulated during tumor evolution. The genomic diversity of primary and metastatic tumors showed an increase over

time. Additionally, our integrative multi-omics analysis adds further evidence that *Rock1* is a crucial player driving metastasis formation, and may be upregulated due to mutations in the PI3K/AKT pathway. In summary, our study contributes towards an improved understanding of melanoma which may eventually lead to improved treatment strategies.

Introduction

Metastatic melanoma is the most lethal type of skin cancer¹. The main contributor to melanoma development is exposure of the skin to UV radiation in sunlight¹. The incidence rate of melanoma in the United States and Europe has been rising over the past decades^{2,3}. The estimated number of new melanoma cases and deaths in 2018 in the United States is 91,270 and 9,320, respectively². The five-year survival rate of metastasized melanoma in the United States is estimated to be around 20%². Progress in understanding the molecular underpinnings of this disease, as well as earlier detection has led to improved survival rates^{1,3}. However, resistance emerges to both oncogene-targeting agents as well immunotherapies, which warrants the need for an improved understanding of tumor progression^{1,4-6}.

Extensive research has led to the identification of several oncogenes, tumor suppressors and key pathways in melanoma that are frequently altered and present potential drug targets⁷. Most commonly, melanomas are driven by oncogenic MAPK signaling driven by activating mutations in *BRAF* or *NRAS*⁸⁻¹⁰. Additionally, the inactivation of the tumor suppressor gene *PTEN* is a frequent event in melanomas^{8,11}, which activates the PI3K/AKT pathway and contributes to tumor invasion¹. Furthermore, Wnt signaling is often activated and contributes to cell proliferation¹.

Dankort and colleagues introduced a mouse model of tamoxifen-inducible, *Braf*-driven and *Pten*-deficient melanoma that mimics the human disease⁸. It offers the possibility to analyze melanoma development and to characterize the genomic and transcriptomic alterations that accompany the mutations in *Braf* and *Pten*. For instance, it has been shown that *Ctnnb1*, which encodes for Catenin beta and belongs to the Wnt pathway, is another key player that is significantly mutated in melanoma¹² and, when activated, drives tumor development together with *Braf* and *Pten*¹³. This mouse model allows to study melanoma progression and metastasis without the large number of potential passenger mutations that may occur in UV-damaged skin and is usually observed in melanoma¹⁴.

Here, we apply the mouse model of *Pten*-deactivated and *Braf*-activated melanomas to study melanoma progression and metastasis in detail. In seven different mice, melanoma was induced on the mouse tail and the tumors were grown over different time periods in each

mouse. As a result, the collected samples are from seven different time points of days after tumor initiation (Figure 1A, upper panel). From each mouse, a primary melanoma sample was collected. Additionally, from each but one mouse, one or two samples from metastases in the lymph nodes were extracted. One normal tail sample served as control. All samples were subjected to whole-exome sequencing (WES), and RNA sequencing (RNA-seq). The samples were collected based on time after tamoxifen-driven tumor induction and not based on tumor size. Hence, the experimental design that includes the timed samples allows the examination of the genomic and transcriptomic changes of melanoma progression over time (Figure 1a, lower panel). Additionally, the comparison of metastatic and primary tumor sites enables the investigation of tumor diversity over time. Moreover, we combined genomic and transcriptomic alterations through a network-based approach (Figure 1b) in order to interpret the set of molecular alterations from a systems perspective. Integrative network analysis can aid to understand the complex interactions between different molecules and to identify novel strategies to combat the disease^{15,16}.

Results

The collected samples are from seven different time points of tumor growth, namely 48, 68, 72, 75, 90, 105, and 143 days (Figure 1a, upper panel; Suppl. Table S1). The samples are referred to by the number of days of tumor growth, as well as the abbreviations “Tu” for primary tumor sample, and “Met” for metastasis samples (Suppl. Table S1). For instance, the primary tumor sample of the first mouse at 48 days of tumor growth is referred to by “48_Tu”. In the following, we will first provide an overview of the detected genomic changes. Next, the timed primary samples are in focus to highlight the transcriptomic changes over time, and network analysis is used to detect important signaling hubs that link genomic alterations to differentially expressed genes. Finally, the genomic and transcriptomic diversity is examined in detail to identify important aspects of metastasis formation.

Overview of genomic changes

The average coverage in the WES samples was 65x (Suppl. Table S1). The median number of single-nucleotide variants (SNVs) and indels that were detected per sample was 25 (Materials and Methods). Overall, the number of genes affected by SNVs and indels was slightly lower for earlier samples, and higher for later samples (Figure 2a). However, this can be attributed mostly to two rather late metastasis samples that seem to be hypermutators

(Figure 2a, upper panel; Suppl. Figure S1). These samples harbored many more mutations than all other samples, which affect known cancer genes (Figure 2a, lower panel). Among the mutated genes in the hypermutators, *Brca2*, *Pik3r1*, *Bclaf1*, *Chek2*, and *Rad51b* play a role in DNA repair and response to DNA damage¹⁷⁻²¹. This suggests that the DNA repair mechanism was defect leading to an accumulation of numerous SNVs and indels²². Moreover, the two hypermutators 75_Met2, and 105_Met have, with 45 and 13, the highest number of deleterious mutations compared to the other samples, which all have less than 10 variants classified as deleterious by SIFT²³.

In general, several of the mutated genes identified here have been associated with melanoma. *Nf1*, which was detected in three samples with a deleterious mutation²³, is a known tumor suppressor gene and among the most commonly mutated genes in melanomas²⁴. It was reported that melanomas with *NF1* mutations are linked to a higher mutational burden and shorter survival¹². Furthermore, in combination with *BRAF* mutations, the inactivation of *NF1* has been shown to be a resistance mechanism against *BRAF* inhibitors²⁵. Additionally, the identified genes *Crebbp*, *Ptprd*, *Map3k1* and *Bclaf1* are recurrently mutated in melanoma, and *Crebbp* was suggested to be a potential driver gene¹⁴. Also, mutations in *RAS* genes, such as *KRAS*, as well as mutations in the PI3K pathway, for instance, in *PIK3R1*, are frequently observed in melanoma¹². By contrast, *LASP1* is mutated in less than 1% of TCGA melanoma samples²⁶, but in our data, *Lasp1* is affected by a deleterious missense mutation in five of 16 samples.

The analysis of copy number variants (CNVs) (Materials and Methods) identified on average five copy number events per sample. These affected on average 345 genes per sample (Figure 2b, upper panel), and all copy number events were copy number gains. Among the affected genes, 14 are in the candidate cancer gene database (Figure 2b, lower panel), which comprises genes that are potential cancer driver genes in mice²⁷. These include two genes encoding for Zinc finger proteins, *Zfp942* and *Zfp40*, which play a role in the regulation of transcription, as well as two genes involved in G-protein coupled receptor activity, namely *Vmn2r40* and *Vmn2r99*²⁸.

Analysis of transcriptomic changes in the primary tumor samples over time

The experimental design allows analyzing the transcriptomic changes over time. First, the timed primary tumor samples were examined in detail. The principal component analysis (PCA) of the primary tumor samples and the normal sample revealed that the tumor samples become more different from the normal with increasing numbers of days of tumor growth (Figure 3a). In particular, the first two tumor samples, 48_Tu, and 68_Tu, are much closer to the normal than the later samples. The PCA of the primary tumor samples without the normal sample is similar to Figure 3a (Suppl. Figure S2), also supporting the clear indication that primary tumors over time become more different from the first time point. In order to detect differentially expressed genes, two comparisons were made with the primary samples and the normal sample. That is, in one comparison, all primary tumor samples were compared against the normal sample (“primary versus normal”), and in the other comparison, the timed primary tumor samples were divided into early and late to identify genes that become up- or downregulated during melanoma development (“primary late versus early”). The cutoff to divide the primary tumor samples into early versus late was 70 days of tumor growth, because this represented the major divide in the PCA (Figure 3a). Hence the early primary tumor samples are 48_Tu and 68_Tu, and the late primary tumor samples are 72_Tu, 75_Tu, 90_Tu, 105_Tu and 143_Tu.

Differential gene expression analysis identified 857 deregulated genes when comparing all seven primary tumor samples to the normal sample (Figure 3b). The comparison of the five late primary samples versus the two early primary tumor samples resulted in 4,320 differentially expressed genes. A total of 546 genes are shared between these two comparisons (Figure 3b), and they are all upregulated genes that generally increase in expression level with increasing numbers of days of tumor growth. In order to identify the functional roles of the deregulated genes, overrepresentation analysis was performed. Among the differentially expressed genes of the comparison “primary late versus early”, those in pathways related to extracellular matrix organization were overrepresented and these genes were mostly upregulated (Figure 3c; Suppl. Figure S3). Degrading and remodeling of the extracellular matrix is important to create a tumor microenvironment that is permissive for invasion and metastasis^{29,30}. The upregulation of genes involved in the extracellular matrix organization has been reported previously in melanoma and was linked to resistance to anti-PD-1 immunotherapy³¹. The overrepresented pathways among the deregulated genes of the “primary versus normal” comparison are in general similar to the ones from the “primary late versus early” comparison (Suppl. Figure S4), but the cell-extracellular matrix interactions seem to be a phenomenon that emerges rather late during tumor development (Figure 3c; Suppl. Figure S3).

Network-based integration of genomic and transcriptomic changes over time

In order to jointly investigate the genomic and transcriptomic changes over time, the tool NetICS (Network-based Integration of Multi-omics Data)¹⁵ was applied. It integrates the genomic and transcriptomic changes through diffusion in a directed functional protein interaction network¹⁵. The underlying rationale is that differential gene expression may be explained by upstream mutated genes. NetICS diffuses the information from genomic alterations forward in the network, and transcriptomic changes are diffused backwards¹⁵. In this way, NetICS identifies major signaling hubs, so-called mediator genes that are relatively close in the network to upstream genomic alterations and downstream differentially expressed genes, and can be regarded as signal transducers that may propagate perturbations from mutations to differentially expressed genes¹⁵. Thereby, NetICS also tries to explain different genomic changes in the different samples that converge through a mediator to the same expression change. Here, the differentially expressed genes from the “primary versus normal” comparison were used, as well as genomic aberrations from all primary tumor samples. As a result, 42 mediator genes were identified (Suppl. Table S2) and ranked across all samples¹⁵. Among these, 27 genes were found in the candidate cancer gene database²⁷, including the gene *Vegfc*. This gene was among the most highly ranked mediator genes, while additionally being highly significant (adjusted p-value < 10⁻³, permutation test). Therefore, *Vegfc* appears to be an important signaling hub between the mutated genes and the differentially expressed genes. The protein interaction network of *Vegfc* was analyzed including the direct downstream genes of *Vegfc* that are differentially expressed, as well as upstream genes with genomic alterations that are up to four levels upwards of *Vegfc* (Figure 4).

In order to gain insight into the timing of these alterations and to determine which parts of the network are perturbed from the early beginning of neoplastic growth, and which changes occur later during the course of melanoma development, the genomic and transcriptomic alterations in this signaling network were examined in more detail. Pertaining to the genomic alterations, with the exception of the copy number gain in *Crlf2*, all mutations in the upstream genes of this *Vegfc* network occurred only in the late samples (Figure 5a, 5b). On the transcriptomic level, most changes occur when comparing the late to the early samples, too. More precisely, all but two of the significantly differentially expressed genes in this network occurred additionally, or exclusively, when comparing expression levels of late versus early primary tumor samples (Figure 5c). In particular, the expression level of *Vegfc* was significantly upregulated in both comparisons “primary versus normal” and “primary late versus early” (Figure 5c) indicating that the expression level of *Vegfc* is upregulated in all primary tumor samples, and that this trend

continues during tumor evolution. This is in line with previous studies demonstrating that the expression level of *VEGFC* increases with melanoma progression³². More precisely, Goydos and Gorski examined *VEGFC* expression levels in 54 human melanoma samples from different phases of progression. They found that *VEGFC* expression was significantly higher in samples from later stages of progression³². Melanoma metastasizes to the lymph nodes through the lymphatic vessels³². *VEGFC* is a growth factor that stimulates lymphangiogenesis and hence is an important factor that enables metastasis through the lymphatic system³³. It has been shown that overexpression of *VEGFC* is a predictor of the occurrence of lymph node metastases in melanoma^{34,35} and that it contributes to immune evasion³⁶.

The upstream mutations of the *Vegfc* network include well known melanoma genes such as *Ctnnb1* and *Nf1*, as well as other, less common genes. *TCF7L2* encodes for TCF4, is part of the Wnt signaling pathway and is known to form a complex with *CTNNB1*^{37,38}. Mutations in this gene have been frequently observed in colorectal cancer³⁹, whereas in melanoma, only 3% of TCGA samples show a *TCF7L2* mutation²⁶. *NTRK2* is mutated in 5% of TCGA melanoma samples²⁶. The missense mutation in *Ntrk2* detected here was classified as deleterious according to the SIFT prediction²³. It was detected in sample 72_Tu, but can also be observed in 90_Tu, and 105_Tu with a variant allele frequency (VAF) of 8% and 11%, respectively, when considering all sequence reads at that position (Figure 5a). In the transcriptome, the expression level of *Ntrk2* is most downregulated in the three samples that show this mutation (Figure 5c), suggesting that the mutation is the cause of the downregulation. The known melanoma gene *Nf1* was affected by two deleterious missense mutations according to the SIFT prediction²³ that were detected in the sample 105_Tu, but the samples 75_Tu, 90_Tu, and 143_Tu also show evidence in the sequence reads for these mutations with VAFs between 2-8% (Figure 5a).

Downstream of *Vegfc*, the differential gene expression analysis identified two growth factor receptors, *Egfr* and *Fgfr2* that are being downregulated in the course of tumor progression (Figure 5c). Two genes, *Kit* and *Met*, are upregulated in all primary tumor samples as compared to the normal, but their gene expression levels show a reversing trend in the latest tumor samples, which show decreased expression after 90 and 100 days of tumor growth (Figure 5c). *Kit* and *MET* are both known oncogenes that are overexpressed in some tumors^{40,41}. *Pdgfra* and *Pdgfrb* are also upregulated in all primary tumor samples and their expression level increases in the late samples (Figure 5c). Overexpression of these two growth factor receptors has been documented for melanomas and other cancer types and was associated with cell proliferation and angiogenesis^{42,43}. We also found that the receptor of *Vegfc*, *Flt4*, which is directly downstream of *Vegfc* in the network (Figure 5c), is increasingly upregulated during melanoma progression. Su and colleagues reviewed how *VEGFC* and its

receptor Flt4, contribute to cancer progression, by, for instance, regulating invasion and cell proliferation⁴⁴. Also, the expression level of VEGFC and Flt4 was linked to cancer progression in cervical cancer ⁴⁵ and prostate cancer ⁴⁶.

Overall, we observe that many of the genes in this *Vegfc* network are members of the MAPK signaling pathway, including *Vegfc* itself⁴⁷. Precisely, besides *Vegfc*, eight of the upstream mutated or differentially expressed genes (*Ptprr*, *Ntrk2*, *Nf1*, *Akt1*, *Akt3*, *Mapk8*, *Mapk14*, *Mapk3*), and all of the 13 downstream deregulated genes are in the MAPK signaling pathway⁴⁷. Our analysis demonstrated that most of these genomic and transcriptomic changes occur later during tumor development, or become stronger in the course of tumor evolution, as, for instance, the upregulation of *Vegfc*. This is in line with a recent melanoma study that also reported accumulating mutations and increased activity of the MAPK signaling pathway during tumor evolution⁴⁸⁻⁵⁰.

Genomic diversity

For all but the first mouse, one or two metastasis samples from the lymph nodes have been extracted in addition to the primary tumor, allowing for the analysis of diversity within each mouse. Here, we examine the genomic diversity with respect to the time measured in days of tumor growth. In order to quantify the diversity and to compare it between the mice, we consider the measure of diversity introduced by Maley and colleagues⁵¹ and inspired by the taxonomic distinctness index in ecology⁵². It was originally defined as the fraction of loci that differ by LOH between two samples⁵¹. In order to make it independent of the number of samples from a neoplasm, Maley et al. used the mean pairwise genetic divergence in case there were more than two samples of a neoplasm⁵¹. They also showed that this measure is positively correlated with probability of progression from neoplasm to adenocarcinoma⁵¹. Here, we adopt the measure suggested by Maley and colleagues and consider instead of numbers of loci with LOH the numbers of genes that are affected by genomic alteration. Hence, we define the mean pairwise diversity score as

$$D = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{X_{ij}}{Y_{ij}},$$

where n is the number of samples per individual mouse, X_{ij} is the number of genes that are affected by SNVs, indels, or copy number change in either sample i or j (but not both), and Y_{ij}

is the total number of genes affected in both sample i and j . In other words, D is the mean fraction of privately mutated genes between all pairs of samples per mouse. The number of samples n per mouse can be different, but the mean pairwise diversity score is comparable between the mice, since the score is normalized by the number of sample pairs per mouse.

This measure was computed for all mice for which two or three samples were available. The analysis revealed that the diversity tends to increase with time of tumor evolution (Figure 6). The regression showed a significant short term increase in diversity between 68 and 105 days of tumor growth (p-value = 0.013; Suppl. Figure S5). The diversity score for the mouse at 75 days of tumor growth was reduced from 0.685 to 0.507 when excluding the hypermutator 75_Met2 (Suppl. Figure S5). However, the regression when excluding both hypermutators (75_Met2, 105_Met) still showed a significant short term increase in diversity between 68 and 90 days of tumor growth (p-value = 0.074; Suppl. Figure S5).

Additionally, phylogenetic trees can be used to interpret genomic diversity since they reveal the different subclones in tumors, including their genotypes and their evolutionary relationship⁵³. This can be of crucial clinical importance, since, for instance, it has been shown that the order of mutations that occurred in tumors can influence factors such as treatment response⁵⁴. Here, the phylogenetic trees were inferred using Cloe⁵⁵ and most trees were linear (Suppl. Figure S6). This finding is in line with those of McGranahan and Swanton, who summarized the phylogenetic trees of different cancer types, and showed that evolutionary trees from melanoma are mostly linear with the bulk of the mutations being clonal⁵⁶.

Analysis of transcriptomic changes between the metastasis and the primary tumor samples

In order to identify which genes or pathways are deregulated in the metastasis samples as compared to the primary samples or the normal, the RNA-seq data from all samples were examined. PCA revealed that the primary tumor samples and metastasis samples differ quite substantially on the transcriptomic level (Figure 7a), albeit part of these differences may be attributable to the fact that the metastasis samples were extracted from the lymph nodes, whereas the primary tumor samples were from the tail. The comparison of all nine metastatic samples versus the normal sample ("metastasis versus normal") resulted in 11,001 deregulated genes (Figure 7b). Differential gene expression analysis between all nine metastases and the 7 primary tumor samples ("metastasis versus primary tumor samples")

revealed 14,999 differentially expressed genes. A total of 8,900 genes are shared between both comparisons. Overall, half of the genes are upregulated, and the other half downregulated. Pathway overrepresentation analysis was performed to map the functions of the deregulated genes. Compared to the primary tumor samples, the genes involved in extracellular matrix organization are not upregulated any more in the metastases (Figure 7c; Suppl. Figure S3; Suppl. Figure S7). Instead, the deregulation of genes in nonsense-mediated decay is significant, and 90% of genes in this process are upregulated (Figure 7c). This can also be observed when comparing the metastasis samples to the normal sample (Suppl. Figure S7). The overexpression of genes involved in nonsense-mediated decay has been reported in cancer^{57,58}. It is a process by which mutated, and also intact, transcripts are being degraded, and hence fulfils a crucial function in regulating the expression levels of other genes^{57,58}. Furthermore, genes involved in keratinization are significantly downregulated compared to the primary tumor samples (Figure 7c). Keratins have been associated with cancer cell invasion and metastasis⁵⁹. While metastasis samples do resemble their tissue of origin⁶⁰, these aforementioned differences between the metastasis and primary tumor samples should be interpreted with caution, as they could also stem from tissue-related differences.

Network-based integration of genomic and transcriptomic changes between the metastasis and primary tumor samples

As before, NetICS was used to combine the mutations and differential gene expression patterns of primary and metastatic samples through diffusion in a directed functional protein interaction network¹⁵. Here, all differentially expressed genes from the “metastasis versus primary tumor samples” were considered, as well as mutated genes that have been detected in either the primary, or the metastatic samples. NetICS identified four significant mediator genes (Suppl. Table S3), of which three (*Tsc1*, *Rock1*, *Spred1*) are also candidate cancer genes²⁷. *Rock1* (Rho-associated protein kinase 1) was selected for a more detailed analysis of its direct network, because it was highly significant (adjusted p-value < 10⁻³, permutation test). This network analysis included the direct downstream genes of *Rock1* that are differentially expressed, as well as upstream genes with genomic alterations that are up to four levels upwards from *Rock1* (Figure 8). Many of the upstream mutations were detected in the two metastasis samples with very high mutational burden, but also four other samples contribute to the set of upstream mutated genes (Figure 8).

Rock1 was overexpressed in the metastasis samples when compared to the primary tumor samples and also in comparison to the normal sample. The overexpression of *ROCK1* has been implicated in metastasis and cell migration⁶¹⁻⁶⁵. In particular, Huang and colleagues demonstrated in a mouse model that the upregulation of *Rock1* is important for metastasis formation in melanoma⁶³. In addition, they reported that the activation of the PI3K/AKT pathway can lead to upregulation of *Rock1* in melanoma⁶³. Among the upstream mutated genes of *Rock1*, nine genes (*Angpt2*, *Col1a2*, *Il7r*, *Kras*, *Lamb1*, *Lamb3*, *Lama4*, *Ntrk2*, *Pik3r1*) are members of the PI3K/AKT signaling pathway⁴⁷. *ROCK1* was also found to be overexpressed and correlated with progression in human bladder cancer⁶⁶, testicular cancer⁶⁷ and human renal cell carcinoma⁶⁸. Lochhead and colleagues reported the detection of activating mutations in *ROCK1* in human cancer that promote tumor progression⁶⁹. In human colorectal cancer, *ROCK1* was significantly upregulated and was associated to poor prognosis⁷⁰.

Discussion

Our integrative multi-omics analysis of timed melanoma samples identified several interesting aspects of melanoma development. First, we found many mutations in known cancer genes including several melanoma-related genes, as well as genes, such as *Lasp1*, that were previously not reported as prominent in melanoma according to TCGA²⁶. *Lasp1* is involved in rearranging the actin cytoskeleton⁷¹, and *Lasp1* knockout studies indicated that loss of *Lasp1* promotes migration of cells and tumor development⁷¹. Second, our analysis confirmed that genes involved in remodeling of the extracellular matrix are upregulated in primary melanoma samples, and additionally revealed that this overexpression increases as the tumor progresses over time. Third, the genes *Vegfc* and *Rock1* were identified as important signaling hubs during melanoma progression and metastasis. Fourth, the analysis of genomic heterogeneity between primary tumor samples and metastases revealed that the diversity within mice increases over time even in the absence of extrinsic mutagenic sources such as UV radiation. Fifth, the nonsense-mediated decay process was identified as enriched among the deregulated genes in the metastasis samples when compared to the primary tumor samples and to the normal sample.

We performed network-based integration of genomic and transcriptomic changes to identify important mediator genes that link mutations to downstream gene deregulation. Cancer therapies may be most efficient if they are targeting entire signaling pathways^{14,16}. The identification of signaling hubs is a crucial step towards this direction¹⁴, and our analysis of important mediator genes identified candidate signaling hubs that could be promising points of

attack. Among the most striking mediator genes in the analysis of the timed primary samples was the lymphangiogenic growth factor *VEGFC*³³, whose direct network revealed many altered genes from the MAPK signaling pathway. *Vegfc* was found to be overexpressed in the primary samples and even more so in the late primary tumor samples. This confirms previous findings that implicate *VEGFC* and lymphatic vessel development in melanoma progression and metastasis^{32,34-36}. Additionally, the network integration highlighted several mutated genes upstream of *Vegfc* that may contribute to its overexpression. In particular, each of the late samples showed evidence in the sequence reads for a subset of the *Vegfc* upstream mutations, which demonstrates that these are mutations that occur during the course of melanoma development. Moreover, since this subset of mutations is different between the late samples implying genomic diversity between mice, the network analysis underlined that these genes are all upstream of the same mediator gene *Vegfc* and that genomic diversity may converge to similar transcriptomic changes. Most of the differentially expressed genes in this *Vegfc* network also show stronger deregulation in the later tumor samples, which provides further evidence of sequential increase in MAPK signaling during melanoma evolution.

The analysis of genomic heterogeneity between primary tumor samples and metastases revealed that the diversity within mice increases over time. This increase is significant when considering the time period between 68 and 105 days of tumor evolution. The last mouse, which is from the time point of 143 days after tumor initiation, demonstrated reduced diversity compared to most other mice from earlier time points (Figure 6; Suppl. Figure S5). The other mice from the time points between 68 and 105 days of tumor growth are at most 15 days apart each, whereas the tumor in the very last mouse evolved for another 38 days more than the one before. This could explain the difference in diversity observed here. However, we conjecture that a larger sample size would be necessary to decide whether the very last mouse is an outlier, or whether the diversity is, in fact, reduced after longer time periods of tumor evolution, for instance after more than 140 days.

In general, the increasing diversity over time observed here, though emphasizes that treatment of tumors becomes increasingly difficult, as the tumor progresses over time. Heterogeneous tumors may harbor many different clones that could be resistant to treatment, and could be undetected when using just a single biopsy^{72,73}.

The inference of phylogenetic trees resulted in phylogenies that are mostly linear, which is in line with other melanoma studies⁵⁶. We also observe that in most mice, at least one third of the total mutations in the tree are already acquired in the first clone of the tree (Suppl. Figure S6). However, since the average coverage was limited to 65x in our data, the phylogenetic trees should be interpreted with caution. The reliability of the tree inference depends on the depth of coverage⁵⁵.

The differential gene expression analysis of metastasis and primary melanoma samples highlighted the importance of the nonsense-mediated decay process in the metastases, which was significantly overrepresented and upregulated in the metastasis samples. Furthermore, the network-based integration confirmed the prominent role of overexpressed *Rock1* in metastasis formation and cell migration. In addition, the network revealed that many of the upstream mutated genes are members of the PI3K/AKT pathway⁴⁷ and may be contributing to *Rock1* overexpression. Shi et al. have previously reported on the preferential acquisition of mutations upregulating the PI3K/AKT pathway in clinical biopsies of progressive melanomas⁴⁹. Our data confirm these findings in a genetically defined mouse model and suggest that this process of PI3K/AKT upregulation is driven by UV-independent mechanisms. We note that many of the mutations in the *Rock1* network were detected in the hypermutator samples. In the samples without a mutation in the depicted network, *Rock1* upregulation could be caused by epigenetic alterations or other alterations further upstream of *Rock1*. *ROCK1* is important for organizing the actin cytoskeleton and is therefore involved in contraction, adhesion and migration of cells, and contributes to metastasis and invasion^{65,74}. The upregulation of *ROCK1* has been observed in several cancer types^{63,74} and has even been associated with decreased survival⁷⁰. Drugs that target *ROCK1* have been proposed^{75,76}, but their effectiveness may also depend on the mutations present in the cells^{74,76}. A novel drug that targets *ROCK1* has been recently proposed for hepatoblastoma⁷⁷, which may also be potent in melanoma. Our study adds further evidence that *Rock1* is a promising drug target that could be exploited to reduce the activity of invasive and metastatic cancer cells. It also highlights mutated genes upstream of *Rock1* that could play a role in *Rock1* deregulation.

We note that our study, which relies on the Braf-activated and Pten-deficient mice, focuses on a subset of melanomas. The induced alterations may influence the genomic and transcriptomic changes observed in the tumors. For instance, *CDKN2A* is commonly lost in human melanoma²⁶, but it does not occur in our samples, and could be due to our mouse model⁷⁸. Future studies with different mouse models or human samples are needed to explore the full spectrum of melanoma. However, the joined occurrence of activated Braf and lost Pten is observed in approximately every fifth human melanoma⁸, and therefore highly relevant.

In summary, our comprehensive analysis of melanoma progression and metastasis formation revealed key genes and pathways involved in these processes, which contributes towards an improved understanding of this disease.

Materials and Methods

Samples. We employed a mouse model of tamoxifen-inducible Tyr::CreERT2, LSL-*Braf*^{V600E}, *Pten*^{fl/fl} melanoma⁸ to obtain samples from seven time points of days of tumor growth. The LSL-*Braf*^{V600E} mice have been received from the Marais Lab⁷⁹. The animal experiments were approved by the Cantonal Veterinary Office Zurich and carried out in accordance to the relevant regulations. Tumor growth was induced on the mouse tail by topical application of 5mg/ml tamoxifen in 100% ethanol. In total we collected seven primary tumor samples and nine samples from metastases in the lymph nodes. Lymph node metastases were not observed in all lymph nodes. The collected metastasis samples were from the inguinal lymph nodes. As a control, a normal tail sample was also collected (Suppl. Table S1).

Exome analysis. The exomes of the samples were sequenced using the Illumina HiSeq 2500 system to generate 101-basepair paired-end reads. For the computational analysis of the data, the NGS-pipe framework⁸⁰ was customized to include the following steps: Adapters were clipped and low-quality bases trimmed using Trimmomatic⁸¹. The reads were aligned to the mouse reference genome version mm10 with bwa⁸². Subsequent processing of the alignment files such as merging, PCR duplicate removal and indexing included the use of the SAMtools⁸³ and Picard tools⁸⁴ suites. Local realignment around indels and base quality recalibration was done using the Genome Analysis Toolkit (GATK)⁸⁵⁻⁸⁷. The depth of coverage was assessed with Qualimap⁸⁸. For the detection of single-nucleotide variants (SNVs), the rank-combination⁸⁹ of deepSNV⁹⁰, JointSNVMix⁹¹, MuTect⁹², SiNVICT⁹³, Strelka⁹⁴, and VarScan2⁹⁵ was employed. Multiple testing correction of the deepSNV p-values was carried out with the R package IHW⁹⁶. Indels were called using the rank-combination of SiNVICT, Strelka, VarDict⁹⁷, and VarScan2. Variants were annotated with the tools SnpSift⁹⁸ and SnpEff⁹⁹. The dbSNP build 142, SNP release v5 from the mouse genome project was used to filter out mutations that are likely not cancerous¹⁰⁰. Copy number variant (CNV) calling was performed with Excavator¹⁰¹ allowing CNVs of length greater or equal than 1Mb, and probability of call greater or equal to 99%. The gonosomes were excluded from the CNV and differential expression analysis since gender differences between the mice may lead to false conclusions. The Variant Effect Predictor¹⁰² was used to obtain SIFT scores for non-synonymous SNVs in coding regions to determine which mutations are deleterious²³. The candidate cancer gene database²⁷ comprises genes that are potential cancer driver genes in mice. Additionally, it contains the information for each gene whether it is listed in the COSMIC database¹⁰³, and whether it is among the Cancer Gene Census genes¹⁰⁴. Throughout this study, we referred to this candidate cancer gene database to judge on the importance of genes. In order to get an overview of the functions of specific genes or proteins, the UniProt database was used¹⁰⁵. The phylogenetic trees were obtained by running ClonE⁵⁵. The parameter for rho, which is the reversion probability, was set to 0.01.

Cloe was run 20 different times with different seeds, and then the tree with highest log-posterior probability was selected for visualization and further inspection.

Transcriptome analysis. RNA-sequencing was done with the Illumina HiSeq 2500 system and 126-basepair single-end reads were generated. Similar as to the exome sequencing analysis, the NGS-pipe framework was employed to analyze the data. The analysis steps included adapter clipping and trimming of low-quality bases using Trimmomatic. Alignment of the reads to the mouse reference genome version mm10 was performed with the STAR aligner¹⁰⁶. The tool featureCounts¹⁰⁷ was used to obtain read counts, and differentially expressed genes were called with DESeq2¹⁰⁸ using a cutoff for adjusted p-values of 0.01. The pathway overrepresentation analysis was performed with WebGestalt¹⁰⁹ and using all expressed genes as a background gene list that had at least a count of 10 fragments across all samples¹¹⁰. The cutoff for the adjusted p-value of an overrepresented pathway was 0.05. Four comparisons were made on the RNA level to detect differentially expressed genes: comparison between all primary tumor samples versus the normal, between late and early primary tumor samples, between metastasis samples and the normal, as well as between the metastasis and primary tumor samples (Suppl. Figure S3).

Combined analysis. For the computational analysis and visualization of the WES and RNA-seq data, several R packages¹¹¹ were used including ggplot2¹¹², reshape2¹¹³, RColorBrewer¹¹⁴, dplyr¹¹⁵, gplots¹¹⁶, treemapify¹¹⁷, VennDiagram¹¹⁸, scales¹¹⁹, cowplot¹²⁰, ComplexHeatmap¹²¹, biomaRt^{122,123}, OrganismDbi¹²⁴, Mus.musculus¹²⁵, igraph¹²⁶, graph¹²⁷, and Rgraphviz¹²⁸. The criterion for selecting the mediator genes Vegfc and Rock1 was the following: Select those genes from the mediator list that are in the candidate cancer gene database²⁷. Among those, choose the gene that is most significant and highly ranked. For Rock1: From the four genes in the list of mediators, Tsc1, Rock1 and Spred1 were in the candidate cancer gene database. Rock1 and Spred1 were highly significant (adjusted p-value < 10⁻³), and Rock1 was higher ranked than Spred1. For Vegfc: The five most highly ranked genes in the list of mediators were all in the candidate cancer gene database. Vegfc was the most significant mediator of the highly ranked genes (adjusted p-value < 10⁻³).

Funding

Part of this work was supported by SystemsX.ch IPhD Grant SXPHI0_142005, by the Novartis Foundation for Medical-Biological Research (Grant 13A25), and by ERC Synergy Grant 609883.

Author contributions

NB and WK designed the study and acquired the funding. WK, SFM, LF provided the mouse samples. CB prepared and supervised the sequencing. ALM conducted the computational sequencing data analysis and visualization. CD ran the NetICS analysis. ALM inspected and visualized results. ALM drafted the manuscript. All authors read and approved the final manuscript.

Availability of data and material

The generated WES and RNA-seq data from the 17 samples is made available at the European Nucleotide Archive under accession number PRJEB28651.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to thank Hans-Joachim Ruscheweyh for helpful advice concerning the differential gene expression analysis.

References

- 1 Bertolotto, C. Melanoma: from melanocyte to genetic alterations and clinical options. *Scientifica (Cairo)* **2013**, 635203, doi:10.1155/2013/635203 (2013).
- 2 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J Clin* **68**, 7-30, doi:10.3322/caac.21442 (2018).
- 3 MacKie, R. M., Hauschild, A. & Eggermont, A. M. Epidemiology of invasive cutaneous melanoma. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **20 Suppl 6**, vi1-7, doi:10.1093/annonc/mdp252 (2009).
- 4 Ratnikov, B. I., Scott, D. A., Osterman, A. L., Smith, J. W. & Ronai, Z. A. Metabolic rewiring in melanoma. *Oncogene* **36**, 147-157, doi:10.1038/onc.2016.198 (2017).
- 5 Jaiswal, B. S. *et al.* ERK Mutations and Amplification Confer Resistance to ERK-Inhibitor Therapy. *Clin Cancer Res*, doi:10.1158/1078-0432.CCR-17-3674 (2018).
- 6 Zamarin, D. *et al.* PD-L1 in tumor microenvironment mediates resistance to oncolytic immunotherapy. *J Clin Invest* **128**, 1413-1428, doi:10.1172/JCI98047 (2018).
- 7 Flaherty, K. T., Hodi, F. S. & Fisher, D. E. From genes to drugs: targeted strategies for melanoma. *Nat Rev Cancer* **12**, 349-361, doi:10.1038/nrc3218 (2012).
- 8 Dankort, D. *et al.* Braf(V600E) cooperates with Pten loss to induce metastatic melanoma. *Nature genetics* **41**, 544-552, doi:10.1038/ng.356 (2009).
- 9 Pollock, P. M. *et al.* High frequency of BRAF mutations in nevi. *Nature genetics* **33**, 19-20, doi:10.1038/ng1054 (2003).
- 10 Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954, doi:10.1038/nature00766 (2002).
- 11 Lin, W. M. *et al.* Modeling genomic diversity and tumor dependency in malignant melanoma. *Cancer Res* **68**, 664-673, doi:10.1158/0008-5472.CAN-07-2615 (2008).
- 12 Cirenajwis, H. *et al.* NF1-mutated melanoma tumors harbor distinct clinical and biological characteristics. *Mol Oncol* **11**, 438-451, doi:10.1002/1878-0261.12050 (2017).

- 13 Damsky, W. E. *et al.* beta-catenin signaling controls metastasis in Braf-activated Pten-deficient melanomas. *Cancer Cell* **20**, 741-754, doi:10.1016/j.ccr.2011.10.030 (2011).
- 14 Mann, M. B. *et al.* Transposon mutagenesis identifies genetic drivers of Braf(V600E) melanoma. *Nature genetics* **47**, 486-495, doi:10.1038/ng.3275 (2015).
- 15 Dimitrakopoulos, C. *et al.* Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics*, doi:10.1093/bioinformatics/bty148 (2018).
- 16 Azmi, A. S., Wang, Z., Philip, P. A., Mohammad, R. M. & Sarkar, F. H. Proof of concept: network and systems biology approaches aid in the discovery of potent anticancer drug combinations. *Mol Cancer Ther* **9**, 3137-3144, doi:10.1158/1535-7163.MCT-10-0642 (2010).
- 17 Havre, P. A., Rice, M. C., Noe, M. & Kmiec, E. B. The human REC2/RAD51B gene acts as a DNA damage sensor by inducing G1 delay and hypersensitivity to ultraviolet irradiation. *Cancer Res* **58**, 4733-4739 (1998).
- 18 Patel, K. J. *et al.* Involvement of Brca2 in DNA repair. *Mol Cell* **1**, 347-357 (1998).
- 19 Bartek, J. & Lukas, J. Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell* **3**, 421-429 (2003).
- 20 Song, L. *et al.* p85alpha acts as a novel signal transducer for mediation of cellular apoptotic response to UV radiation. *Mol Cell Biol* **27**, 2713-2731, doi:10.1128/MCB.00657-06 (2007).
- 21 Lee, Y. Y., Yu, Y. B., Gunawardena, H. P., Xie, L. & Chen, X. BCLAF1 is a radiation-induced H2AX-interacting partner involved in gammaH2AX-mediated regulation of apoptosis and DNA repair. *Cell Death Dis* **3**, e359, doi:10.1038/cddis.2012.76 (2012).
- 22 Chae, Y. K. *et al.* Genomic landscape of DNA repair genes in cancer. *Oncotarget* **7**, 23312-23321, doi:10.18632/oncotarget.8196 (2016).
- 23 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-1081, doi:10.1038/nprot.2009.86 (2009).
- 24 Krauthammer, M. *et al.* Exome sequencing identifies recurrent mutations in NF1 and RASopathy genes in sun-exposed melanomas. *Nature genetics* **47**, 996-1002, doi:10.1038/ng.3361 (2015).
- 25 Maertens, O. *et al.* Elucidating distinct roles for NF1 in melanomagenesis. *Cancer Discov* **3**, 338-349, doi:10.1158/2159-8290.CD-12-0313 (2013).
- 26 Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *The New England journal of medicine* **375**, 1109-1112, doi:10.1056/NEJMp1607591 (2016).
- 27 Abbott, K. L. *et al.* The Candidate Cancer Gene Database: a database of cancer driver genes from forward genetic screens in mice. *Nucleic acids research* **43**, D844-848, doi:10.1093/nar/gku770 (2015).
- 28 Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045-3046, doi:10.1093/bioinformatics/btp536 (2009).
- 29 Kessenbrock, K., Plaks, V. & Werb, Z. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **141**, 52-67, doi:10.1016/j.cell.2010.03.015 (2010).
- 30 Botti, G. *et al.* Microenvironment and tumor progression of melanoma: new therapeutic perspectives. *J Immunotoxicol* **10**, 235-252, doi:10.3109/1547691X.2012.723767 (2013).
- 31 Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35-44, doi:10.1016/j.cell.2016.02.065 (2016).
- 32 Goydos, J. S. & Gorski, D. H. Vascular endothelial growth factor C mRNA expression correlates with stage of progression in patients with melanoma. *Clin Cancer Res* **9**, 5962-5967 (2003).
- 33 Stacker, S. A., Achen, M. G., Jussila, L., Baldwin, M. E. & Alitalo, K. Lymphangiogenesis and cancer metastasis. *Nat Rev Cancer* **2**, 573-583, doi:10.1038/nrc863 (2002).
- 34 Boone, B. *et al.* The role of VEGF-C staining in predicting regional metastasis in melanoma. *Virchows Arch* **453**, 257-265, doi:10.1007/s00428-008-0641-6 (2008).
- 35 Spirc, Z., Eri, Z. & Eric, M. Lymphatic vessel density and VEGF-C expression as independent predictors of melanoma metastases. *J Plast Reconstr Aesthet Surg* **70**, 1653-1659, doi:10.1016/j.bjps.2017.06.040 (2017).

- 36 Lund, A. W. *et al.* VEGF-C promotes immune tolerance in B16 melanomas and cross-presentation of tumor antigen by lymph node lymphatics. *Cell Rep* **1**, 191-199, doi:10.1016/j.celrep.2012.01.005 (2012).
- 37 Rosenbluh, J., Wang, X. & Hahn, W. C. Genomic insights into WNT/beta-catenin signaling. *Trends Pharmacol Sci* **35**, 103-109, doi:10.1016/j.tips.2013.11.007 (2014).
- 38 Roose, J. & Clevers, H. TCF transcription factors: molecular switches in carcinogenesis. *Biochimica et biophysica acta* **1424**, M23-37 (1999).
- 39 Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337, doi:10.1038/nature11252 (2012).
- 40 Abbaspour Babaei, M., Kamalidehghan, B., Saleem, M., Huri, H. Z. & Ahmadipour, F. Receptor tyrosine kinase (c-Kit) inhibitors: a potential therapeutic target in cancer cells. *Drug Des Devel Ther* **10**, 2443-2459, doi:10.2147/DDDT.S89114 (2016).
- 41 Comoglio, P. M., Trusolino, L. & Boccaccio, C. Known and novel roles of the MET oncogene in cancer: a coherent approach to targeted therapy. *Nat Rev Cancer* **18**, 341-358, doi:10.1038/s41568-018-0002-y (2018).
- 42 Barnhill, R. L., Xiao, M., Graves, D. & Antoniades, H. N. Expression of platelet-derived growth factor (PDGF)-A, PDGF-B and the PDGF-alpha receptor, but not the PDGF-beta receptor, in human malignant melanoma in vivo. *Br J Dermatol* **135**, 898-904 (1996).
- 43 Raica, M. & Cimpean, A. M. Platelet-Derived Growth Factor (PDGF)/PDGF Receptors (PDGFR) Axis as Target for Antitumor and Antiangiogenic Therapy. *Pharmaceuticals (Basel)* **3**, 572-599, doi:10.3390/ph3030572 (2010).
- 44 Su, J. L. *et al.* The role of the VEGF-C/VEGFR-3 axis in cancer progression. *Br J Cancer* **96**, 541-545, doi:10.1038/sj.bjc.6603487 (2007).
- 45 Van Trappen, P. O. *et al.* Expression of vascular endothelial growth factor (VEGF)-C and VEGF-D, and their receptor VEGFR-3, during different stages of cervical carcinogenesis. *J Pathol* **201**, 544-554, doi:10.1002/path.1467 (2003).
- 46 Jennbacken, K., Vallbo, C., Wang, W. & Damber, J. E. Expression of vascular endothelial growth factor C (VEGF-C) and VEGF receptor-3 in human prostate cancer is associated with regional lymph node metastasis. *Prostate* **65**, 110-116, doi:10.1002/pros.20276 (2005).
- 47 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research* **44**, D457-462, doi:10.1093/nar/gkv1070 (2016).
- 48 Shain, A. H. *et al.* Genomic and Transcriptomic Analysis Reveals Incremental Disruption of Key Signaling Pathways during Melanoma Evolution. *Cancer Cell* **34**, 45-55 e44, doi:10.1016/j.ccell.2018.06.005 (2018).
- 49 Shi, H. *et al.* Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discov* **4**, 80-93, doi:10.1158/2159-8290.CD-13-0642 (2014).
- 50 Johnson, D. B. *et al.* Acquired BRAF inhibitor resistance: A multicenter meta-analysis of the spectrum and frequencies, clinical behaviour, and phenotypic associations of resistance mechanisms. *Eur J Cancer* **51**, 2792-2799, doi:10.1016/j.ejca.2015.08.022 (2015).
- 51 Maley, C. C. *et al.* Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature genetics* **38**, 468-473, doi:10.1038/ng1768 (2006).
- 52 R., C. K. & M., W. R. A taxonomic distinctness index and its statistical properties. *Journal of Applied Ecology* **35**, 523-531, doi:10.1046/j.1365-2664.1998.3540523.x (1998).
- 53 Beerewinkel, N., Schwarz, R. F., Gerstung, M. & Markowetz, F. Cancer evolution: mathematical models and computational inference. *Syst Biol* **64**, e1-25, doi:10.1093/sysbio/syu081 (2015).
- 54 Ortmann, C. A. *et al.* Effect of mutation order on myeloproliferative neoplasms. *The New England journal of medicine* **372**, 601-612, doi:10.1056/NEJMoa1412098 (2015).
- 55 Marass, F., Mouliere, F., Yuan, K., Rosenfeld, N. & Markowetz, F. A phylogenetic latent feature model for clonal deconvolution. *The Annals of Applied Statistics* **10**, 2377-2404 %@ 1932-6157 (2016).

- 56 McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613-628, doi:10.1016/j.cell.2017.01.018 (2017).
- 57 Gardner, L. B. Nonsense-mediated RNA decay regulation by cellular stress: implications for tumorigenesis. *Mol Cancer Res* **8**, 295-308, doi:10.1158/1541-7786.MCR-09-0502 (2010).
- 58 Nickless, A., Bailis, J. M. & You, Z. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell Biosci* **7**, 26, doi:10.1186/s13578-017-0153-7 (2017).
- 59 Karantza, V. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene* **30**, 127-138, doi:10.1038/onc.2010.456 (2011).
- 60 Suzuki, M. & Tarin, D. Gene expression profiling of human lymph node metastases and matched primary breast carcinomas: clinical implications. *Mol Oncol* **1**, 172-180, doi:10.1016/j.molonc.2007.03.005 (2007).
- 61 Sun, J. *et al.* Targeting the metastasis suppressor, NDRG1, using novel iron chelators: regulation of stress fiber-mediated tumor cell migration via modulation of the ROCK1/pMLC2 signaling pathway. *Mol Pharmacol* **83**, 454-469, doi:10.1124/mol.112.083097 (2013).
- 62 Gilkes, D. M. *et al.* Hypoxia-inducible factors mediate coordinated RhoA-ROCK1 expression and signaling in breast cancer cells. *Proc Natl Acad Sci U S A* **111**, E384-393, doi:10.1073/pnas.1321510111 (2014).
- 63 Huang, G. X. *et al.* Up-regulation of Rho-associated kinase 1/2 by glucocorticoids promotes migration, invasion and metastasis of melanoma. *Cancer Lett* **410**, 1-11, doi:10.1016/j.canlet.2017.09.005 (2017).
- 64 Maskey, N. *et al.* MicroRNA-340 inhibits invasion and metastasis by downregulating ROCK1 in breast cancer cells. *Oncol Lett* **14**, 2261-2267, doi:10.3892/ol.2017.6439 (2017).
- 65 Itoh, K. *et al.* An essential part for Rho-associated kinase in the transcellular invasion of tumor cells. *Nat Med* **5**, 221-225, doi:10.1038/5587 (1999).
- 66 Kamai, T. *et al.* Significant association of Rho/ROCK pathway with invasion and metastasis of bladder cancer. *Clin Cancer Res* **9**, 2632-2641 (2003).
- 67 Kamai, T. *et al.* Overexpression of RhoA, Rac1, and Cdc42 GTPases is associated with progression in testicular cancer. *Clin Cancer Res* **10**, 4799-4805, doi:10.1158/1078-0432.CCR-0436-03 (2004).
- 68 Abe, H. *et al.* Possible role of the RhoC/ROCK pathway in progression of clear cell renal cell carcinoma. *Biomed Res* **29**, 155-161 (2008).
- 69 Lochhead, P. A., Wickman, G., Mezna, M. & Olson, M. F. Activating ROCK1 somatic mutations in human cancer. *Oncogene* **29**, 2591-2598, doi:10.1038/onc.2010.3 (2010).
- 70 Li, J., Bharadwaj, S. S., Guzman, G., Vishnubhotla, R. & Glover, S. C. ROCK I Has More Accurate Prognostic Value than MET in Predicting Patient Survival in Colorectal Cancer. *Anticancer Res* **35**, 3267-3273 (2015).
- 71 Zhang, H. *et al.* Lasp1 gene disruption is linked to enhanced cell migration and tumor formation. *Physiol Genomics* **38**, 372-385, doi:10.1152/physiolgenomics.00048.2009 (2009).
- 72 Yates, L. R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* **21**, 751-759, doi:10.1038/nm.3886 (2015).
- 73 Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature genetics* **46**, 225-233, doi:10.1038/ng.2891 (2014).
- 74 Wei, L., Surma, M., Shi, S., Lambert-Cheatham, N. & Shi, J. Novel Insights into the Roles of Rho Kinase in Cancer. *Arch Immunol Ther Exp (Warsz)* **64**, 259-278, doi:10.1007/s00005-015-0382-6 (2016).
- 75 Sadok, A. *et al.* Rho kinase inhibitors block melanoma cell migration and inhibit metastasis. *Cancer Res* **75**, 2272-2284, doi:10.1158/0008-5472.CAN-14-2156 (2015).
- 76 Vogel, C. J. *et al.* Cooperative induction of apoptosis in NRAS mutant melanoma by inhibition of MEK and ROCK. *Pigment Cell Melanoma Res* **28**, 307-317, doi:10.1111/pcmr.12364 (2015).

- 77 Liu, W. *et al.* Lycorine inhibits cell proliferation and migration by inhibiting ROCK1/cofilin-induced actin dynamics in HepG2 hepatoblastoma cells. *Oncol Rep* **40**, 2298-2306, doi:10.3892/or.2018.6609 (2018).
- 78 Daniotti, M. *et al.* BRAF alterations are associated with complex mutational profiles in malignant melanoma. *Oncogene* **23**, 5968-5977, doi:10.1038/sj.onc.1207780 (2004).
- 79 Dhomen, N. *et al.* Oncogenic Braf induces melanocyte senescence and melanoma in mice. *Cancer Cell* **15**, 294-303, doi:10.1016/j.ccr.2009.02.022 (2009).
- 80 Singer, J. *et al.* NGS-pipe: a flexible, easily extendable, and highly configurable framework for NGS analysis. *Bioinformatics*, doi:10.1093/bioinformatics/btx540 (2017).
- 81 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170 (2014).
- 82 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 83 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 84 Picard. *Picard*. Accessed 31 August 2017., <<http://broadinstitute.github.io/picard/>> (
- 85 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
- 86 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491-498, doi:10.1038/ng.806 (2011).
- 87 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33, doi:10.1002/0471250953.bi1110s43 (2013).
- 88 Okonechnikov, K., Conesa, A. & Garcia-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292-294, doi:10.1093/bioinformatics/btv566 (2016).
- 89 Hofmann, A. L. *et al.* Detailed simulation of cancer exome sequencing data reveals differences and common limitations of variant callers. *BMC Bioinformatics* **18**, 8, doi:10.1186/s12859-016-1417-7 (2017).
- 90 Gerstung, M. *et al.* Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nature communications* **3**, 811, doi:10.1038/ncomms1814 (2012).
- 91 Roth, A. *et al.* JointSNVMix: a probabilistic model for accurate detection of somatic mutations in normal/tumour paired next-generation sequencing data. *Bioinformatics* **28**, 907-913, doi:10.1093/bioinformatics/bts053 (2012).
- 92 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology* **31**, 213-219 (2013).
- 93 Kockan, C. *et al.* SiNVICT: ultra-sensitive detection of single nucleotide variants and indels in circulating tumour DNA. *Bioinformatics* **33**, 26-34, doi:10.1093/bioinformatics/btw536 (2017).
- 94 Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811-1817, doi:10.1093/bioinformatics/bts271 (2012).
- 95 Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research* **22**, 568-576 (2012).
- 96 Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods* **13**, 577-580, doi:10.1038/nmeth.3885 (2016).
- 97 Lai, Z. *et al.* VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic acids research* **44**, e108, doi:10.1093/nar/gkw227 (2016).
- 98 Cingolani, P. *et al.* Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in genetics* **3** (2012).

99 Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide
polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2;
iso-3. *Fly (Austin)* **6**, 80-92, doi:10.4161/fly.19695 (2012).

100 Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene
regulation. *Nature* **477**, 289-294, doi:10.1038/nature10413 (2011).

101 Magi, A. *et al.* EXCAVATOR: detecting copy number variants from whole-exome sequencing
data. *Genome Biol* **14**, R120, doi:10.1186/gb-2013-14-10-r120 (2013).

102 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122,
doi:10.1186/s13059-016-0974-4 (2016).

103 Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic acids research*
45, D777-D783, doi:10.1093/nar/gkw1121 (2017).

104 Futreal, P. A. *et al.* A census of human cancer genes. *Nat Rev Cancer* **4**, 177-183,
doi:10.1038/nrc1299 (2004).

105 The UniProt, C. UniProt: the universal protein knowledgebase. *Nucleic acids research* **45**,
D158-D169, doi:10.1093/nar/gkw1099 (2017).

106 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
doi:10.1093/bioinformatics/bts635 (2013).

107 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930,
doi:10.1093/bioinformatics/btt656 (2014).

108 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

109 Wang, J., Vasaiakar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more
comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit.
Nucleic acids research **45**, W130-W137, doi:10.1093/nar/gkx356 (2017).

110 Timmons, J. A., Szkop, K. J. & Gallagher, I. J. Multiple sources of bias confound functional
enrichment analysis of global -omics data. *Genome Biol* **16**, 186, doi:10.1186/s13059-015-
0761-7 (2015).

111 Team, R. C. R: A Language and Environment for Statistical Computing. [https://www.R-](https://www.R-project.org/)
[project.org/](https://www.R-project.org/). *R Foundation for Statistical Computing, Vienna, Austria* (2018).

112 Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* (2009).

113 Wickham, H. Reshaping Data with the reshape Package. *Journal of Statistical Software* **21**, 1-
20 (2007).

114 Neuwirth, E. RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. [https://CRAN.R-](https://CRAN.R-project.org/package=RColorBrewer)
[project.org/package=RColorBrewer](https://CRAN.R-project.org/package=RColorBrewer). (2014).

115 Hadley Wickham, R. F., Lionel Henry and Kirill Müller. dplyr: A Grammar of Data
Manipulation. R package version 0.7.4. <https://CRAN.R-project.org/package=dplyr>. (2017).

116 Gregory R. Warnes, B. B., Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy
Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz,
Bill Venables. gplots: Various R programming tools for plotting data. *R package* (2016).

117 Wilkins, D. treemapify: Draw Treemaps in 'ggplot2'. *R package version* (2018).

118 Chen, H. VennDiagram: Generate High-Resolution Venn and Euler Plots. *R package* (2018).

119 Wickham, H. scales: Scale Functions for Visualization. *R package* (2017).

120 Wilke, C. O. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. *R package*
(2017).

121 Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in
multidimensional genomic data. *Bioinformatics* **32**, 2847-2849,
doi:10.1093/bioinformatics/btw313 (2016).

122 Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases
and microarray data analysis. *Bioinformatics* **21**, 3439-3440,
doi:10.1093/bioinformatics/bti525 (2005).

- 123 Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of
genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191,
doi:10.1038/nprot.2009.97 (2009).
- 124 Marc Carlson, H. P., Martin Morgan, Valerie Obenchain. OrganismDbi: Software to enable the
smooth interfacing of different database packages. *R package* (2017).
- 125 Team, B. C. Mus.musculus: Annotation package for the Mus.musculus object. *R package*
(2015).
- 126 Gabor Csardi, T. N. The igraph software package for complex network research. *InterJournal*
Complex Systems, 1695 (2006).
- 127 R. Gentleman, E. W., W. Huber, S. Falcon. graph: graph: A package to handle graph data
structures. *R package* (2017).
- 128 Kasper Daniel Hansen, J. G., Li Long, Robert Gentleman, Seth Falcon, Florian Hahne,
Deepayan Sarkar. Rgraphviz: Provides plotting capabilities for R graph objects. *R package*
(2017).

Figures

Figure 1. (a) Upper panel: Experimental design. The analysis includes a tamoxifen-inducible mouse model of *Braf*^{A600E}-driven, *Pten*-deficient melanoma⁸ from seven time points of days of tumor growth. From each of the seven mice, one primary tumor sample was collected. For all but the first mouse, also one or two lymph node metastasis samples were extracted. As a control, a normal tail sample was collected. **Lower panel:** Schematic representation of the aspects under examination. The genomic and transcriptomic changes over time are investigated by tracking the variant allele frequency or gene expression level in the timed samples. Additionally, the comparison of primary and metastatic tumor sites enables an analysis of the diversity over time. **(b)** For each sample, whole-exome and RNA sequencing was performed, allowing network-based integration of genomic alterations and transcriptomic changes.

Figure 2. Overview of the genomic diversity. **(a)** Upper panel: The numbers of genes that are affected by a single-nucleotide variant (SNV) or indel per sample. The color indicates the annotation of the mutation (Full barplot in Suppl. Figure S1). Lower panel: The heatmap shows all cancer gene census genes that harbor a nonsynonymous mutation in a coding region or a splice site mutation (grey box in the legend 'Alterations'). Multiple mutations in a gene are indicated with the black dot, and the color represents the annotation of highest priority, which is the same ordering as in the legend from top to bottom. For instance, if a gene had a missense and a synonymous mutation, the color will be red, and the black dot indicates that there were additional mutations. In order to clearly show which sample name is referring to which sample type and mouse, the numbers of days of tumor growth as well as the sample type is indicated. **(b)** Upper panel: The number of genes that are affected by a copy number change. The color indicates on which chromosome the copy number change occurred. The gonosomes are omitted from the plot and further analysis (Materials and Methods). All copy number events displayed were copy number gains. Lower panel: The heatmap depicting the genes affected by a copy number gain that are in the candidate cancer gene database²⁷. As before, the numbers of days of tumor growth as well as the sample type is indicated for each sample.

Figure 3. Transcriptome analysis of timed primary tumor samples. **(a)** Principal component analysis (PCA) of the expression levels of the 500 most variable genes when including all primary tumor samples and the normal sample. The number of days of tumor growth is indicated in the plot. **(b)** Differential gene expression analysis revealed 857 deregulated genes when comparing the primary samples to the normal. Most of them are upregulated as shown by the heatmap, where pink indicates higher expression levels, and green lower expression levels. The expression levels visualized here are normalized, and additionally centered and scaled such that each row has mean zero and standard deviation one. Moreover, differential gene expression analysis was performed to compare the five late primary tumor samples (> 70 days) to the two early primary tumor samples (48_Tu, 68_Tu). A total of 4,320 genes was differentially expressed in this comparison, where 546 genes were significantly upregulated in both comparisons. **(c)** Pathway overrepresentation analysis was performed with all 4,320 differentially expressed genes of the late versus early primary tumor samples comparison. The colors represent the top-level pathways of the Reactome hierarchy, and each individual pathway which was significantly overrepresented, is indicated in the treemap. The area of the tiles mirrors the significance level. The rounded adjusted p-value as well as the percentage of upregulated genes among the differentially expressed genes is indicated.

Figure 4. Network of mediator gene *Vegfc*. NetICS was performed to integrate RNA-seq and WES data. The displayed *Vegfc* network includes the upstream genes with non-silent genomic variation in at least one of the samples up to four levels upstream, as well as the direct downstream genes that are differentially expressed. Also, the genes connecting the upstream mutated genes with the mediator are indicated. Here, for reasons of biological interpretation, the deregulation of a gene is displayed if it is differentially expressed in either the “primary versus normal”, or the “primary late versus early” comparison. Additionally, for biological interpretation, also differential expression upstream of *Vegfc* is indicated, even if the upstream deregulated genes did not contribute to the mediator *Vegfc* in NetICS, since deregulation is diffused backwards. Some of the nodes summarize several genes in order to preserve space in the display. In such cases, the edges between nodes indicate that the edge is true for at least one of the genes in the summarized set. The table on the upper right indicates in which samples the mutations were detected.

Figure 5. Genomic and transcriptomic alterations in the *Vegfc* mediator network. (a) Variant allele frequency of the upstream mutations in the primary tumor samples that are highlighted in the mediator network of Figure 4. **(b)** Copy number status of *Crif2*. **(c)** Normalized gene expression levels of the differentially expressed genes in the network of Figure 4. In particular, the genes are separated into different categories depending on whether their deregulation occurred in all primary tumor samples, or only in the late samples. Genes with adjusted p-value < 0.01 were declared as being differentially expressed.

Figure 6. Genomic diversity over time. Venn diagrams displaying the number of genes affected by SNVs, indels or copy number change that are shared or private between the samples from each mice. The diversity score (y axis) is the mean fraction of privately mutated genes between all pairs of samples per mouse. In general, the diversity score tends to increase with the days of tumor growth. This trend is true for all but the last mouse, in which the tumor was grown much longer. The overall trend of increasing genomic diversity over time is still true when excluding the two hypermutators 75_Met2 and 105_Met (Suppl. Figure S5). Moreover, when considering the mice between 68 and 105 days of tumor growth, a significant short term increase in diversity can be observed (Suppl. Figure S5).

Figure 7. Transcriptomes of metastasis samples in comparison to primary tumor samples. **(a)** Principal component analysis (PCA) of expression levels of the 500 most variable genes when including all metastasis and primary tumor samples. The number of days of tumor growth is indicated in the plot. **(b)** Differential gene expression analysis revealed 11,001 deregulated genes when comparing the metastasis samples to the normal. Additionally, differential gene expression analysis was performed to compare the metastasis samples to the primary tumor samples. A total of 14,999 genes was differentially expressed in this comparison, where the 8,900 genes were significantly upregulated in both comparisons. **(c)** Pathway overrepresentation analysis was performed with all differentially expressed genes of the metastasis versus primary comparison. The colors represent the top-level pathways of the Reactome hierarchy, and each individual pathway which was significantly overrepresented, is indicated in the treemap. The area of the tiles mirrors the significance level.

Figure 8. Network of mediator gene *Rock1* in the metastatic versus primary samples comparison. The upstream genes with nonsynonymous genomic variation up to four levels upstream are displayed, as well as the direct downstream genes that are differentially expressed. Also, the genes connecting the upstream mutated genes with the mediator are indicated. In this network, some of the upstream genes are also differentially expressed, but this is not visualized. The table indicates in which samples the mutations were detected.