



ARTICLE

<https://doi.org/10.1038/s41467-018-08069-x>

OPEN

# Enhancer hijacking activates oncogenic transcription factor NR4A3 in acinic cell carcinomas of the salivary glands

Florian Haller <sup>1</sup>, Matthias Bieg<sup>2,3</sup>, Rainer Will<sup>4</sup>, Cindy Körner<sup>5</sup>, Dieter Weichenhan<sup>6</sup>, Alexander Bott<sup>5</sup>, Naveed Ishaque<sup>2,3</sup>, Pavlo Lutsik <sup>6</sup>, Evgeny A. Moskalev<sup>1</sup>, Sarina K. Mueller<sup>7</sup>, Marion Bähr<sup>6</sup>, Angelika Woerner<sup>5</sup>, Birgit Kaiser<sup>4</sup>, Claudia Scherl<sup>7</sup>, Marlen Haderlein<sup>8</sup>, Kortine Kleinheinz<sup>9</sup>, Rainer Fietkau<sup>8</sup>, Heinrich Iro<sup>7</sup>, Roland Eils<sup>2,3,10,11</sup>, Arndt Hartmann<sup>1</sup>, Christoph Plass<sup>6</sup>, Stefan Wiemann<sup>4,5</sup> & Abbas Agaimy<sup>1</sup>

The molecular pathogenesis of salivary gland acinic cell carcinoma (AciCC) is poorly understood. The secretory Ca-binding phosphoprotein (SCPP) gene cluster at 4q13 encodes structurally related phosphoproteins of which some are specifically expressed at high levels in the salivary glands and constitute major components of saliva. Here we report on recurrent rearrangements [t(4;9)(q13;q31)] in AciCC that translocate active enhancer regions from the SCPP gene cluster to the region upstream of *Nuclear Receptor Subfamily 4 Group A Member 3 (NR4A3)* at 9q31. We show that NR4A3 is specifically upregulated in AciCCs, and that active chromatin regions and gene expression signatures in AciCCs are highly correlated with the NR4A3 transcription factor binding motif. Overexpression of NR4A3 in mouse salivary gland cells increases expression of known NR4A3 target genes and has a stimulatory functional effect on cell proliferation. We conclude that NR4A3 is upregulated through enhancer hijacking and has important oncogenic functions in AciCC.

<sup>1</sup>Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Krankenhausstr. 8-10, 91054 Erlangen, Germany. <sup>2</sup>Center for Digital Health, Berlin Institute of Health and Charité – Universitätsmedizin Berlin, Kapelle-Ufer 2, 10117 Berlin, Germany. <sup>3</sup>Heidelberg Center for Personalized Oncology (DKFZ-HIPO), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. <sup>4</sup>Genomics and Proteomics Core Facility, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany. <sup>5</sup>Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany. <sup>6</sup>Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. <sup>7</sup>Department of Otorhinolaryngology, Head & Neck Surgery, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Waldstrasse 1, 91054 Erlangen, Germany. <sup>8</sup>Department of Radiation Therapy, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Universitätsstrasse 27, 91054 Erlangen, Germany. <sup>9</sup>Division of Theoretical Bioinformatics (B080), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. <sup>10</sup>Health Data Science Unit, University Hospital Heidelberg, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany. <sup>11</sup>Translational Lung Research Center Heidelberg, German Center for Lung Research, University of Heidelberg, Im Neuenheimer Feld 156, 69120 Heidelberg, Germany. These authors contributed equally: Florian Haller, Matthias Bieg, Stefan Wiemann, Abbas Agaimy. Correspondence and requests for materials should be addressed to F.H. (email: [florian.haller@uk-erlangen.de](mailto:florian.haller@uk-erlangen.de)) or to S.W. (email: [s.wiemann@dkfz-heidelberg.de](mailto:s.wiemann@dkfz-heidelberg.de))

A group of structurally related genes encoding secretory Ca-binding phosphoproteins (SCPPs) map to a genomic region of ~750 kb on chromosome 4q13, i.e., the SCPP gene cluster<sup>1</sup>. These genes evolved from a common ancestor gene by tandem duplication events<sup>1</sup>. The encoded salivary proteins, milk caseins and enamel matrix proteins are functionally related through their ability to regulate the calcium phosphate levels in the extracellular environment. The function of the salivary proteins includes protection of tooth enamel by stabilizing saliva supersaturated with calcium salts, as well as antimicrobial and antifungal properties<sup>1</sup>. The expression of the majority of the salivary gland genes within the SCPP gene cluster (e.g., *STATH*, *HTN1*, *HTN3*, *ODAM*, *FDCSP*, and *MUC7*) is mostly restricted to the salivary glands, where they rank among the most highly expressed genes (<https://gtexportal.org/home/>).

Acinic cell carcinoma (AcICC) shows serous differentiation closely resembling normal acini of salivary glands, which produce major components of saliva<sup>2</sup>. While no recurrent molecular genetic event in AcICC has been reported yet<sup>2</sup>, several other salivary gland neoplasms are characterized by specific recurrent genomic rearrangements<sup>3,4</sup>. These typically result in the generation of oncogenic chimeric fusion genes, e.g., the *NFIB-MYB* gene fusion in adenoid cystic carcinomas (ACC)<sup>5</sup>. Only recently, an alternate mechanism of oncogenic upregulation of the same transcription factor MYB through genomic rearrangements involving the translocation of enhancer regions has been demonstrated in ACC lacking the *NFIB-MYB* gene fusion<sup>6</sup>. Similarly, “regulatory rearrangements” driving oncogene overexpression through enhancer hijacking or insulator dysfunction are increasingly recognized in other cancer types<sup>7,8</sup>.

For the current study, we employ comprehensive genomic, transcriptomic, and epigenomic profiling of AcICCs and normal parotid gland tissue to decipher the molecular genetic events in AcICC. We show that AcICCs harbor recurrent and specific rearrangements [t(4;9)(q13;q31)] that bring active enhancer regions from the SCPP gene cluster in close proximity to the transcription start site (TSS) of transcription factor *Nuclear Receptor Subfamily 4 Group A Member 3 (NR4A3)* at 9q31. The genomic breakpoints within the SCPP gene cluster correlate with high H3K27ac active chromatin marks that are present in both normal parotid gland and AcICC tumor tissues, and we demonstrate significant enhancer activity within the breakpoint region in a functional analysis. We show that NR4A3 is specifically and consistently upregulated in AcICCs. There is an increase in active chromatin marks and upregulation of gene expression in AcICC tumor tissues compared to normal parotid gland that is significantly correlated with the NR4A3 transcription factor binding motif. Functional analyses of mouse salivary gland and human mammary gland cells with NR4A3 overexpression show significant effects on gene expression including known NR4A3 target genes, as well as stimulatory effects on cell proliferation and cell cycle. We conclude that the genomic rearrangements [t(4;9)(q13;q31)] are the initiating tumorigenic events in AcICCs and that these rearrangements enable upregulation of NR4A3 through inappropriate influence of active chromatin regions derived from the SCPP gene cluster. We suggest that NR4A3 upregulation is an early oncogenic event in AcICCs through constitutive stimulation of cell proliferation.

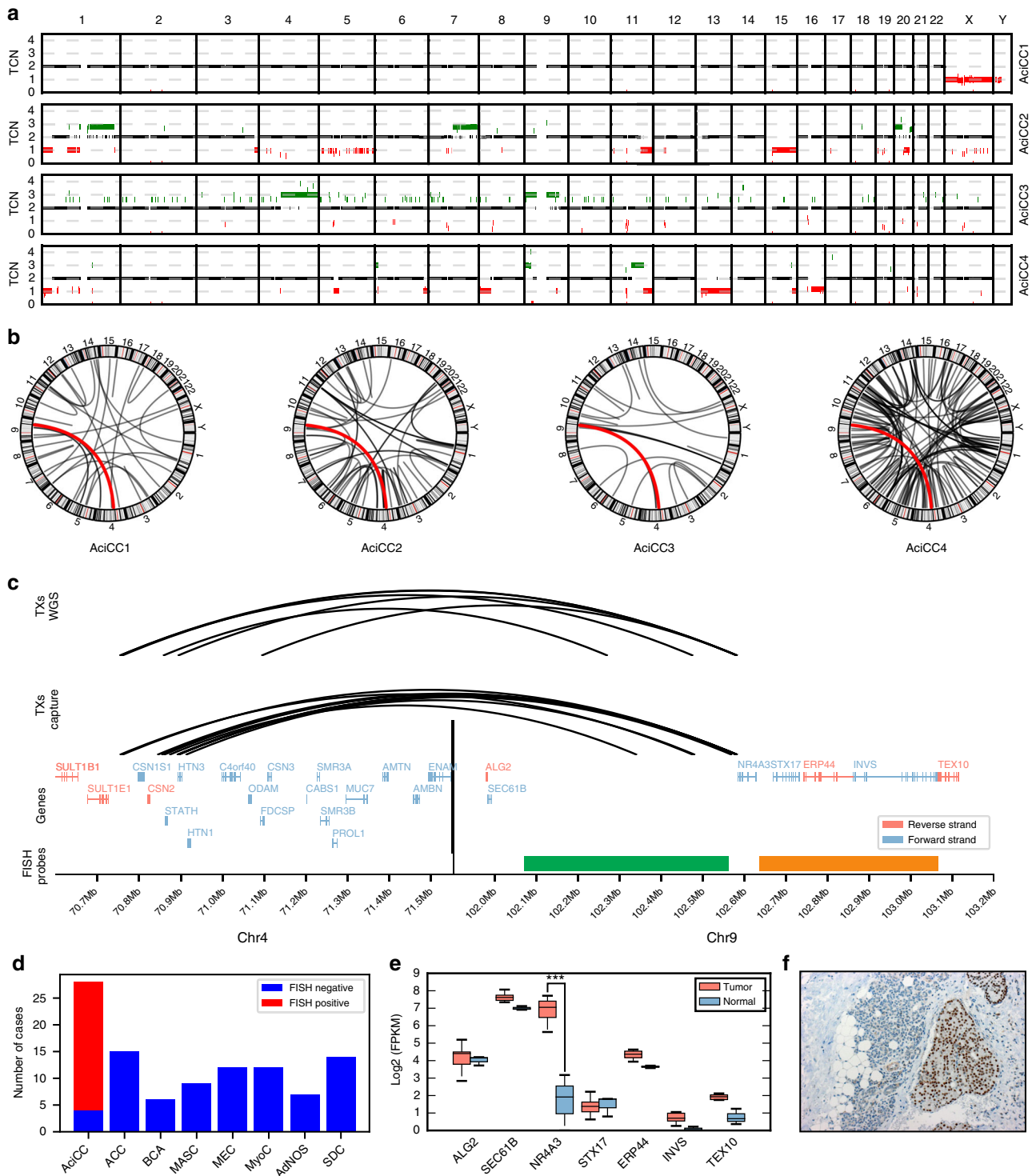
## Results

**Identification of recurrent rearrangements [t(4;9)(q13;q31)].** Whole Genome Sequencing (WGS) of four paired tumor/normal AcICC cases and tumor WGS in two additional AcICCs revealed that copy number changes were few and broadly distributed among the tumor genomes without recurrent gains or losses present in all

cases (Fig. 1a, Supplementary Data 1). The tumors carried very few somatic non-synonymous variants in the exonic fraction (SNVs: mean 14, range 7–17; InDels mean: 3, range 2–7), and no gene was mutated in more than one tumor (Supplementary Table 1). In contrast, a specific inter-chromosomal rearrangement [t(4;9)(q13;q31)] was present in all six AcICCs (Fig. 1b). The 4q13 breakpoints were distributed within ~340 kb at the SCPP gene cluster spanning eight genes (Fig. 1c, upper panel left side). All of the 9q31 breakpoints were located upstream of *NR4A3*, with four events located within a ~600 bp region directly upstream of the TSS (Fig. 1c, upper panel right side). We thus focused on the rearrangement [t(4;9)(q13;q31)] and *NR4A3* as a potential driver in AcICC, and established a hybrid capture based next generation sequencing approach with probes covering the ~315 kb upstream of the *NR4A3* TSS to identify nine additional AcICCs with a [t(4;9)(q13;q31)] translocation (Fig. 1c, middle panel).

**Transcription factor NR4A3 is upregulated in AcICCs.** We further employed Fluorescence in-Situ Hybridization (FISH) using a break-apart probe flanking the *NR4A3* gene locus (Fig. 1c lower panel right side, Supplementary Figure 1) to evaluate the frequency of this translocation event in a total of 29 AcICCs and 75 other salivary gland neoplasms. Split signals indicating a *NR4A3* gene locus rearrangement were present in 24 of 28 evaluable AcICCs (86%), while no *NR4A3* translocation was observed in any of 75 non-AcICC salivary gland tumors (Fig. 1d, Supplementary Table 2). Whole transcriptome RNA-sequencing analysis of ten AcICCs (Supplementary Data 2) revealed no coding chimeric fusion transcripts as described previously<sup>9</sup>. However, as compared to normal parotid salivary gland tissue, significant upregulation of *NR4A3*, the gene closest to the 9q31 breakpoints, was determined in all AcICCs, while there was no significant difference in the expression of other genes located upstream and further downstream of the 9q31 breakpoints (Fig. 1e). Twenty-eight of 29 AcICCs (97%) showed nuclear immunostaining for NR4A3 (Supplementary Table 2), while NR4A3 was not detectable in normal salivary gland tissue (Fig. 1f).

**The SCPP gene cluster harbors highly active chromatin marks.** The lack of a coding gene fusion and the observation that the 4q13 breakpoints were rearranged in either forward or reverse orientation to the *NR4A3* upstream region directed us towards regulatory chromatin marks (Fig. 2, Supplementary Data 3), which function independent of the genomic sequence orientation. We employed Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) using antibodies against core histone modifications H3K27ac (active promoter/enhancer), H3K4me3 (active promoter), H3K27me3 (repressed chromatin), and against the transcription factors NR4A3 and CCCTC-Binding Factor (CTCF), as well as whole genome bisulfite sequencing (Supplementary Data 4), to decipher epigenetic regulatory marks around the breakpoint regions. This data were complemented with published data on topologically associated domains (TADs) and chromatin contacts (HiC)<sup>10</sup>. We observed enrichment of active H3K27ac and H3K4me3 chromatin marks and low DNA methylation levels broadly distributed throughout the 4q13 SCPP gene cluster in normal parotid gland tissue, consistent with abundant expression of the SCPP salivary gland genes mapping there (Fig. 2 left panel). Comparison with published chromatin contact data<sup>10</sup> indicated three sub-TAD domains within a larger TAD, encompassing the majority of the highly expressed salivary gland genes (e.g., *STATH*, *HTN3*, *HTN1*) within a single sub-TAD region characterized by abundant active chromatin marks, with inversely orientated CTCF binding sites correlating well with the borders of the sub-TAD regions. In contrast, there were both



active H3K27ac/H3K4me3 and repressive H3K27me3 chromatin marks at the 9q31 locus upstream of the *NR4A3* TSS and very-low DNA methylation levels at the promoter region in the normal parotid tissue, indicating a bivalent/poised promoter (Fig. 2 right panel). Comparing three AcicC samples with normal parotid tissue, the chromatin marks, as well as DNA methylation patterns were overall well-preserved in the tumors, with slightly lower repressive H3K27me3 chromatin marks at the *NR4A3* promoter region and lower DNA methylation levels at the *NR4A3* gene body in two samples. CHIP-seq for NR4A3 binding sites revealed several peaks at the SPP gene cluster in the tumor tissues that

corresponded well with active chromatin marks, while there was no significant NR4A3 binding signal in the normal parotid tissue. Regarding the *NR4A3* gene locus, there were NR4A3 binding peaks at the upstream region of the *NR4A3* gene locus both in the normal parotid and tumor tissues.

**Three distinct patterns of rearrangements [t(4;9)(q13;q31)].** Notably, in all six AcicCs analyzed by WGS, the 4q13 genomic breakpoints mapped to regions with an enrichment of active chromatin marks associated with abundantly expressed salivary

**Fig. 1** Identification of recurrent rearrangements t(4;9)(q13;q31) in AcICCs. **a** Plots of copy number changes in four AcICCs with paired tumor-normal whole genome sequencing (WGS) show gains (green) and losses (red) in the individual tumors. **b** Circos plots of translocations in the four cases, with recurrent rearrangements t(4;9)(q13;q31) highlighted in red. **c** Upper panel: Detailed mapping of t(4;9)(q13;q31) chromosomal breakpoints of the four AcICCs with paired tumor-normal WGS and two AcICCs with only tumor WGS demonstrate the distribution of 4q13 breakpoints among ~340 kbps spanning eight different genes at the SCPP gene cluster (left side), and clustering of the 9q31 breakpoints upstream of the *NR4A3* gene locus (right side). Middle panel: Detailed mapping of t(4;9)(q13;q31) chromosomal breakpoints of nine additional AcICCs with tumor hybrid capture sequencing data confirms the pattern of 4q13 breakpoints within the SCPP gene cluster (left) and upstream of the *NR4A3* gene locus (right). Lower panel: SCPP gene cluster (left side), *NR4A3* and neighboring genes (right side) with green and orange bars indicating the location of *NR4A3* break apart FISH probes. **d** Absolute number of cases with genomic rearrangements of the *NR4A3* gene locus in 28 AcICCs and 75 other salivary gland neoplasms analyzed by FISH. **e** mRNA expression (log2 FPKM values) of *NR4A3* and neighboring genes in ten tumor samples and three normal parotid gland samples, with only *NR4A3* showing a significant upregulation (deSeq2<sup>47</sup>; \*\*\**P* < 0.001; Box-plot center line: median; bounds of box: 25 and 75% quantiles; whiskers: minimum and maximum values). **f** *NR4A3* immunohistochemistry demonstrating strong nuclear expression in AcICC tumor tissue (right), but absence in normal parotid gland tissue (left). Source data for Figs. 1b, 1c and 1e are provided as a Source Data file. Source Data for Fig. 1d are provided in Supplementary Table 2. TCN, Tumor Copy Number; TXs, Translocations; SCPP, secretory Ca-binding phosphoprotein; AcICC, Acinar cell carcinoma; ACC, adenoid cystic carcinoma; BCA, basal cell adenocarcinoma; MASC, mammary analog secretory carcinoma; MyoC, myoepithelial carcinoma; AdNOS, adenocarcinoma not otherwise specified; SDC, salivary duct carcinoma

gland genes (Fig. 2 left panel). Adding the additional nine AcICCs with rearrangement [t(4;9)(q13;q31)] identified by the hybrid capture assay, three distinct patterns of rearrangements could be established (Fig. 3). In four AcICCs (AcICC5, AcICC4, AcICC3, AcICC19), the 4q13 breakpoint was located in a small genomic region that fell between two inversely oriented CTCF binding sites that were characterized by active H3K27ac peaks and correlated with the border between two sub-TAD regions as indicated by published chromatin contact data (Figs. 2 and 3). In all four samples the breakpoints were orientated in the same direction, e.g., the 4q13 reverse strand was fused to the 9q31 forward strand, so that in all four cases the resulting rearranged sequence included the genomic region of the SCPP gene cluster. The second and most frequent pattern was present in ten AcICCs, which all harbored the 4q13 breakpoint within a genomic region encompassing the three most abundantly expressed salivary gland genes *STATH*, *HTN3*, and *HTN1* with correlating highly active chromatin marks (Fig. 3). In contrast to the tumors with the first pattern, both orientations of rearrangements of the 4q13 genomic region were observed in this group, with three tumors having the 4q13 forward strand fused to the *NR4A3* upstream region, and seven the 4q13 reverse strand. The third pattern was identified in only one sample (AcICC2) where the 4q13 breakpoint fell directly within the first intron of the salivary gland gene *FDCSP*, but no chimeric fusion reads between *FDCSP* and *NR4A3* were detected on the RNA level. Further exploiting the ChIP-seq data, we next modeled the potential interaction of active H3K27ac and H3K4me3 chromatin marks from the 4q13 SCPP gene cluster and the *NR4A3* gene locus as a result of the genomic rearrangements for three individual tumor samples AcICC3, AcICC1 and AcICC2, representing the three different breakpoint patterns, respectively, (Fig. 4). In AcICC3 (Fig. 4b) representing the second most common group with breakpoints between two CTCF binding sites, active chromatin marks from the SCPP gene cluster were situated directly at the breakpoint as well as ~90 kb, ~130 kb and ~290 kb upstream of the breakpoint, with the latter three peaks ranking among the highest 5% of peaks and thus fulfilling the criteria of super-enhancers (Supplemental Figure S5a). In AcICC1 (Fig. 4c) representing the most frequently observed pattern with breakpoints within the genomic region encoding for *STATH*, *HTN3* and *HTN1*, active chromatin marks derived from the 4q13 SCPP gene cluster were situated directly at the breakpoint, as well as ~25 kb upstream of the breakpoint, and these chromatin marks ranked high within the highly active chromatin marks in this sample but did not fulfill the criteria of super-enhancers (Supplemental Figure 2). In AcICC2 (Fig. 4d) representing its own pattern with the breakpoint directly within the *FDCSP* gene,

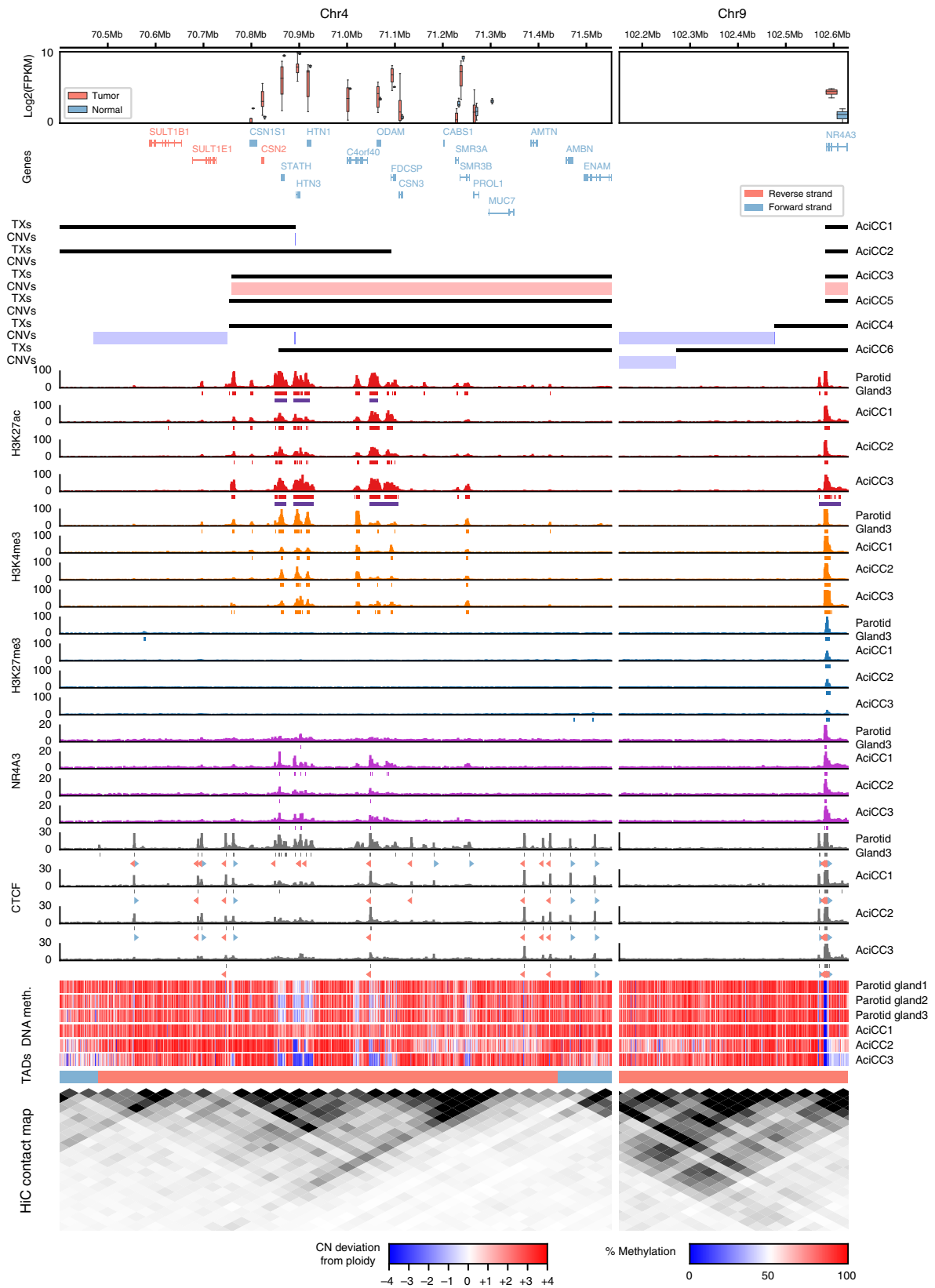
active chromatin marks from the *FDCSP* promoter region were directly located at the breakpoint region, as well as more distal at ~28 kb, ~174 kb, and ~228 kb upstream of the breakpoint. In contrast, no active chromatin marks were found in normal parotid tissue without genomic rearrangement within the same distance upstream of the *NR4A3* TSS (Fig. 4a). In all three tumors analyzed, the active chromatin marks derived from the SCPP gene cluster were partially accompanied by *NR4A3* peaks, and there was a reduction of inhibitory H3K27me3 peaks at the *NR4A3* promoter region. Additionally, a CTCF binding site 13.5 kb upstream of the *NR4A3* gene locus was separated from the ultimate *NR4A3* promoter region in all three AcICCs, possibly disrupting an insulator function.

#### Analysis of enhancer activity within the SCPP gene cluster.

Since no minimal overlapping genomic region within the SCPP gene cluster could be defined among the 15 AcICCs with recurrent [t(4;9)(q13;q31)], we focused on the most common group with breakpoints within the genomic region encoding for *STATH*, *HTN3*, and *HTN1*. Putative enhancer regions were identified by presence of active chromatin marks within this region and comparison with a published data set<sup>11</sup>. We focused on two proximal and distal regions upstream of the *STATH* TSS with highly active chromatin marks in our normal parotid gland ChIP-seq data, including a putative enhancer region predicted by published chromatin segmentation data within the distal region (Fig. 5). Ten fragments with a size of 340–846 bp mapping within these regions were cloned and analyzed for enhancer activity using a dual luciferase reporter assay. Indeed, the fragment mapping to genomic coordinates chr4:70849125–70849971 correlated to the predicted enhancer region<sup>11</sup> and displayed significant enhancer activity in this analysis (two-sided *T*-Test, *P* < 0.01, Fig. 5).

#### Gene expression correlates with *NR4A3* binding motif.

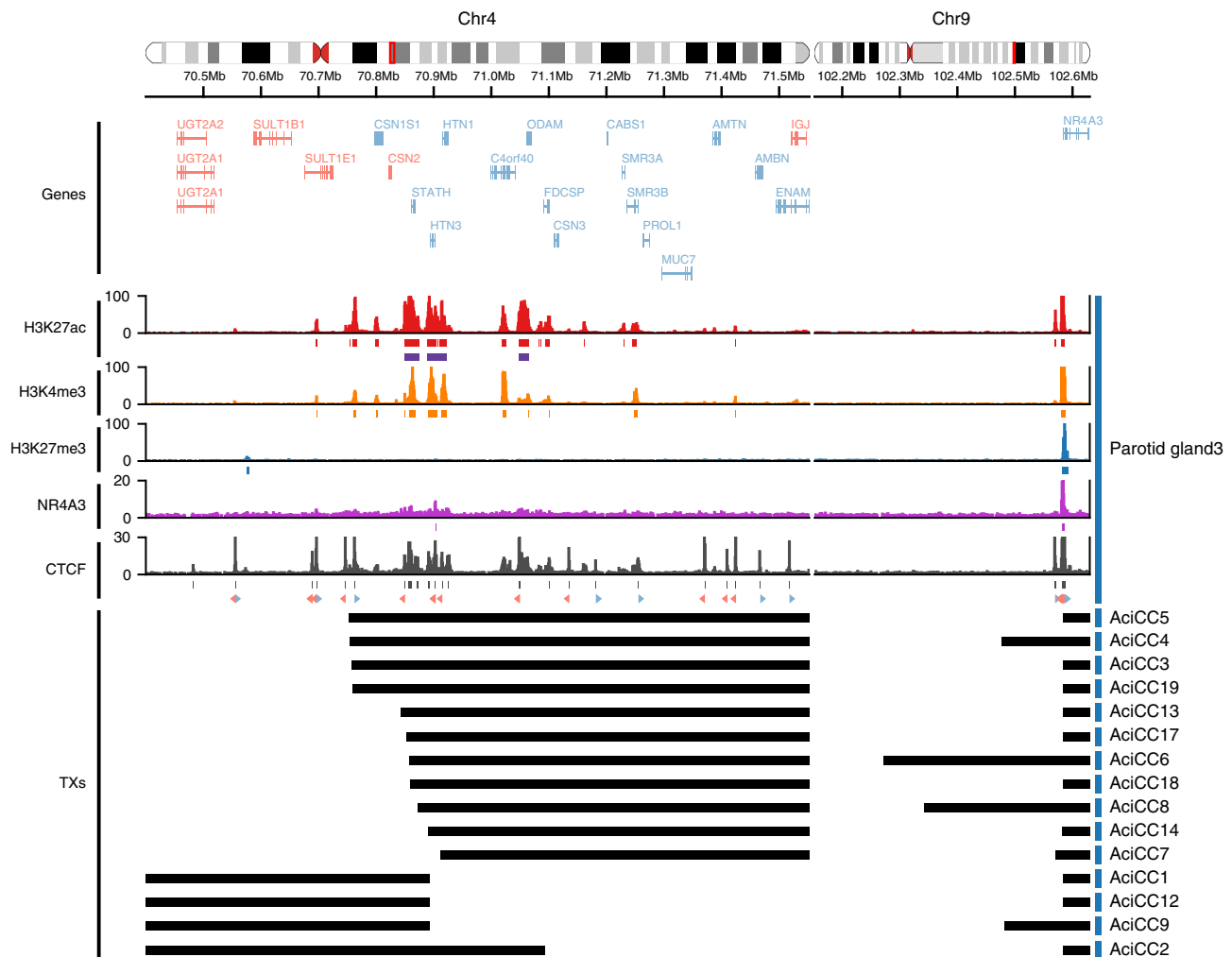
To analyze the correlation between *NR4A3* transcription factor upregulation and global mRNA expression in AcICCs, we compared whole transcriptome RNA sequencing and ChIP-seq data. Gene expression analysis of ten AcICCs and three normal parotid tissues identified 827 and 404 genes that were significantly up- and downregulated in AcICCs vs. normal parotid tissues, respectively (Fig. 6a, Supplementary Data 2). Functional enrichment analysis of genes that were upregulated in the tumors revealed a significant enrichment of genes involved in DNA replication, cell division, and DNA damage response (annotation cluster 1 to 4 and 6), and of genes involved in inflammatory response and complement



activation (annotation cluster 5 and 7). Among the downregulated genes, there was a significant enrichment of genes involved in GTPase activation and salivary secretion (annotation cluster 1 and 2, respectively). A comparison of differentially expressed genes with differential ChIP-seq peaks for active histone mark H3K27ac and NR4A3 transcription factor revealed a positive correlation between

differential H3K27ac and NR4A3 peaks (Supplementary Figure 3), and a significant enrichment of higher H3K27ac and NR4A3 peaks was observed among upregulated genes in the tumors. Approximately 25% and 15% of upregulated genes were associated with upregulated H3K27ac and NR4A3 peaks in the tumors, respectively (Fig. 6b). While ~ 25% of downregulated genes were associated with

**Fig. 2** Genomic breakpoints in AciCCs correlate with active chromatin marks and NR4A3 binding sites at the 4q13 SCPP gene cluster. Summary of (top to bottom) mRNA expression ( $\log_2(\text{FPKM})$ ), gene loci (Genes), genomic translocation breakpoints (TXs), copy-number variations (CNVs), active (H3K27ac, H3K4me3) and repressive (H3K27me3) histone marks, NR4A3 binding sites, CTCF binding sites, DNA methylation, topologically associated domains (TADs), and chromatin contacts (HiC)<sup>10</sup> for the chromosomal regions surrounding the 4q13 (left panel) and 9q31 (right panel) breakpoints in normal parotid gland and AciCC tumor tissues. mRNA expression is shown for ten AciCC tumors and three normal parotid gland samples in red and blue, respectively (Box-plot centre line: median; bounds of box: 25 and 75% quantiles; whiskers: extend to last value greater than  $Q1-1.5 \times \text{IQR}$ , and last value less than  $Q3+1.5 \times \text{IQR}$  respectively). Here IQR is the inter quartile range, Q1 is the first, and Q3 the third quartile). Black bars demonstrating the translocations correspond to genomic material included in the rearrangement, with red and blue bars indicating gains and losses, respectively. For each ChIP-seq experiment, ChIP signals (barcharts), as well as corresponding peaks (directly below ChIP signals) are shown. For H3K27ac, the super-enhancer peaks are shown in addition (purple). Furthermore, CTCF motifs within CTCF peaks are shown (below CTCF peaks). Blue arrows indicate motifs on the forward strand, whereas red arrows indicate motifs on the reverse strand. The DNA methylation tracks show the average methylation within 1 kb binned regions. Different TADs are indicated in different colors

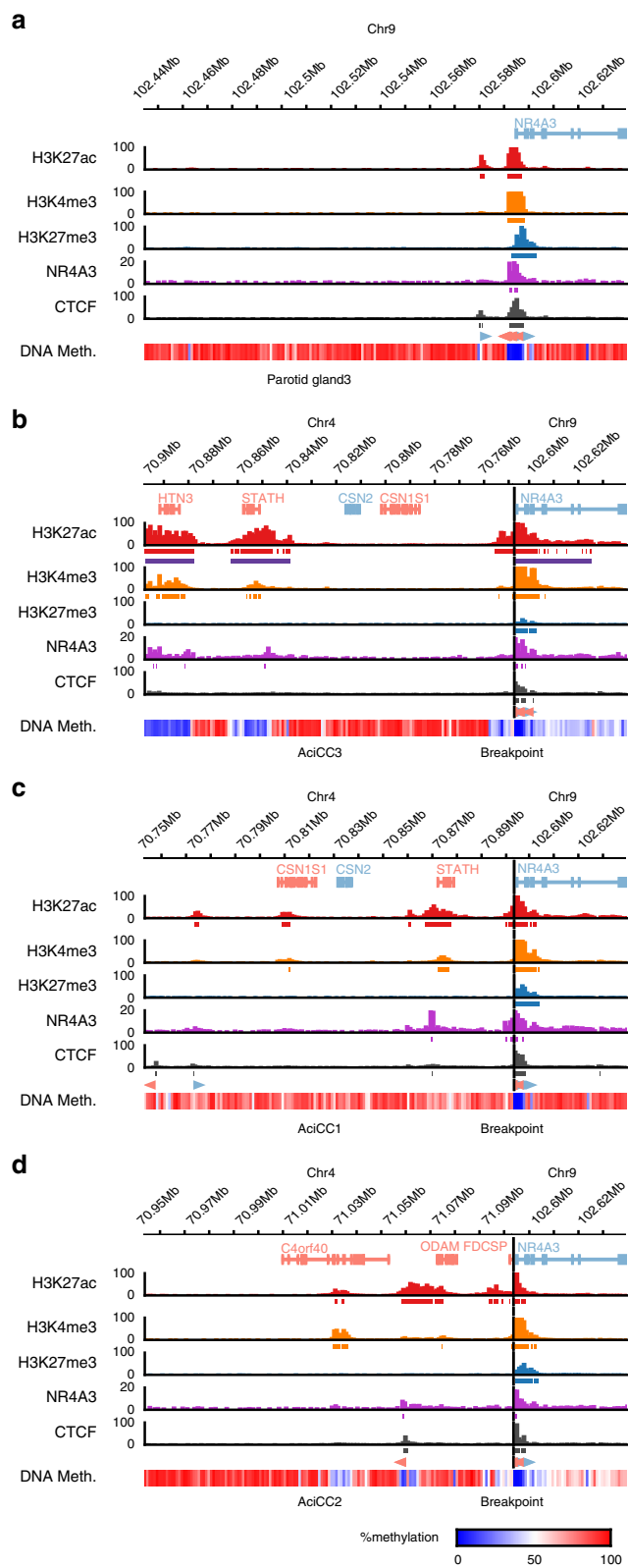


**Fig. 3** Detailed characterization of rearrangements  $t(4;9)(q13;q31)$  in 15 AciCCs identify three different breakpoint patterns. Detailed presentation of rearrangements in 15 AciCCs with  $t(4;9)(q13;q31)$ , with black bars indicating genomic material included in the rearrangement. Active (H3K27ac, H3K4me3) and repressive (H3K27me3) histone marks, NR4A3 binding sites and CTCF binding sites in normal parotid gland tissue are presented for comparison. Source data for translocation breakpoints are provided as a Source Data file

reduced H3K27ac peaks, there were only few reduced NR4A3 peaks in the tumor tissues, and only 2% of the downregulated genes were associated with reduced NR4A3 peaks. Motif analysis of differential H3K27ac peaks in the three tumors vs. the normal parotid tissue identified a significant enrichment of the NGFI-B response element (NBRE) (Fig. 6b), which is the main binding motif for NR4A family members<sup>12</sup>. Correspondingly, de-novo motif analysis of enriched NR4A3 peaks in ChIP-seq data from three AciCC tumor samples revealed the same motif in all three cases (Fig. 6b). The known

NR4A3 target genes *CCND1*<sup>13</sup> and *ENO3*<sup>14</sup> were higher expressed in AciCC tissues on the mRNA level compared to normal parotid gland tissue (Supplementary Figure 4), and this was confirmed on the protein level by immunohistochemistry (Supplementary Figure 5, Supplementary Table 2).

**Functional effects of NR4A3 overexpression in vitro.** To demonstrate the driver activity of NR4A3 in cellular systems, we stably transduced immortalized primary mouse submandibular



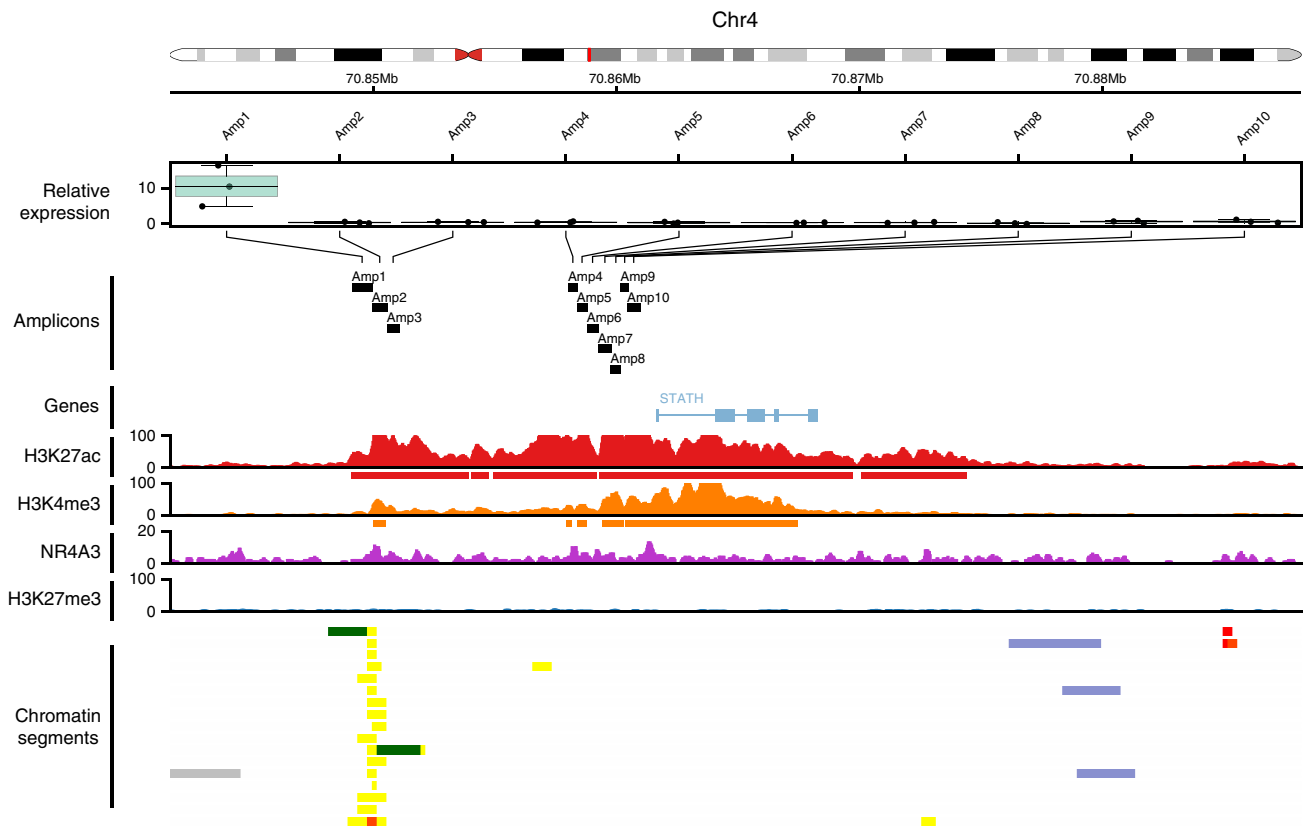
**Fig. 4** Rearrangements t(4;9)(q13;q31) in AcicCs juxtapose the *NR4A3* gene locus proximal to active enhancers and *NR4A3* binding sites. Detailed presentation of active (H3K27ac, H3K4me3) and repressive (H3K27me3) histone marks, *NR4A3* binding sites, CTCF binding sites, and DNA methylation levels in normal parotid gland (**a**) and three distinct AcicC tumor samples AcicC3 (**b**), AcicC1 (**c**) and AcicC2 (**d**). For each ChIP-seq experiment, ChIP signals (barcharts) as well as corresponding peaks (directly below ChIP signals) are shown. For H3K27ac, the super-enhancer peaks are shown in addition (purple). Furthermore, CTCF motifs within CTCF peaks are shown (below CTCF peaks). Blue arrows indicate motifs on the forward strand, whereas red arrows indicate motifs on the reverse strand. The DNA methylation tracks show the average methylation within 1 kb binned regions

*NR4A3* was found expressed only in the *NR4A3* stably transduced cells both at mRNA and protein levels but not in the controls. Mouse *Nr4a3* was not detected in any condition. Mouse *Ccnd1* and *Eno3* mRNAs were found significantly upregulated and mouse *Ccnd1* and *Eno3* proteins were present at higher levels specifically in *NR4A3* overexpression conditions (Fig. 6d). This was validated by qRT-PCR for *Ccnd1* and *Eno3* (Supplementary Figure 8), and by western blot for *Ccnd1* (Supplementary Figure 9). Furthermore, there was a significant upregulation of 432 genes and downregulation of 561 genes in *NR4A3* expressing cells compared to the control cells on the mRNA level (source data file). Nine proteins, including *NR4A3*, were found expressed only in the cells stably transduced with *NR4A3*, while 17 proteins were exclusively detected in the proteomic datasets of controls (Supplementary Table 3). Second, to test for potential functional consequences of *NR4A3* expression in the mouse salivary gland cells, cell viability and proliferation were evaluated. Indeed, expression of *NR4A3* increased the number of cells accumulating with time and also enhanced the fraction of cells in S-phase of the cell cycle (Fig. 6e). To verify that these observations are not cell line or species specific, we stably transduced the human mammary gland cell line MCF10A with the *NR4A3* coding sequence and then selected clones with and without *NR4A3* expression (qRT-PCR) for functional testing. Consistent with the data obtained in the mouse salivary gland cell line model, the MCF10A cells proliferated significantly more when *NR4A3* was expressed and a similar increase in the fraction of cells in S-phase was observed (Fig. 6f).

**Discussion**

*NR4A3* is a member of the *NR4A* subfamily of nuclear receptors that are involved in manifold physiological processes including cell proliferation, differentiation, and metabolism<sup>10</sup>. Through binding to the NBRE motif<sup>10</sup>, *NR4A* proteins are transcription factors that directly modulate gene expression, but can also be part of larger trans-repressive complexes<sup>15,16</sup>. A fusion gene comprised of *NR4A3* and *EWSR1* results from a recurrent translocation [t(9;22)(q31;q12)] in extraskelatal myxoid chondrosarcoma and is the oncogenic driver event in this entity<sup>17</sup>. In contrast, *NR4A3* has been described as tumor suppressor in hematopoietic neoplasms, e.g., acute myeloid leukemia<sup>18</sup>. Likely, the cellular context and the presence or absence of further co-factors determines the effect of *NR4A3* function on gene transcription and phenotypes. Only few genes including the cell cycle regulator *Cyclin D1* (*CCND1*)<sup>13</sup>, the metabolic enzyme *Enolase 3* (*ENO3*)<sup>14</sup>, the secreted serum and extracellular matrix glycoprotein *Vitronectin* (*VTN*)<sup>19</sup>, and the nuclear receptor and transcriptional regulator *Peroxisome Proliferator Activated Receptor Gamma* (*PPARG*)<sup>20</sup> have been functionally validated as direct targets of *NR4A3* or the *NR4A3*-*EWSR1* fusion protein in

gland (SG) cells either with the human *NR4A3* open reading frame (ORF) or with red firefly luciferase coding regions and investigated molecular and phenotypic effects. First, we performed RNA-seq and whole mass spectrometry proteome analysis for unbiased determination of mRNA expression (Supplementary Figure 6, Supplementary Data 5) and protein abundance (Supplementary Figure 7), respectively. Human



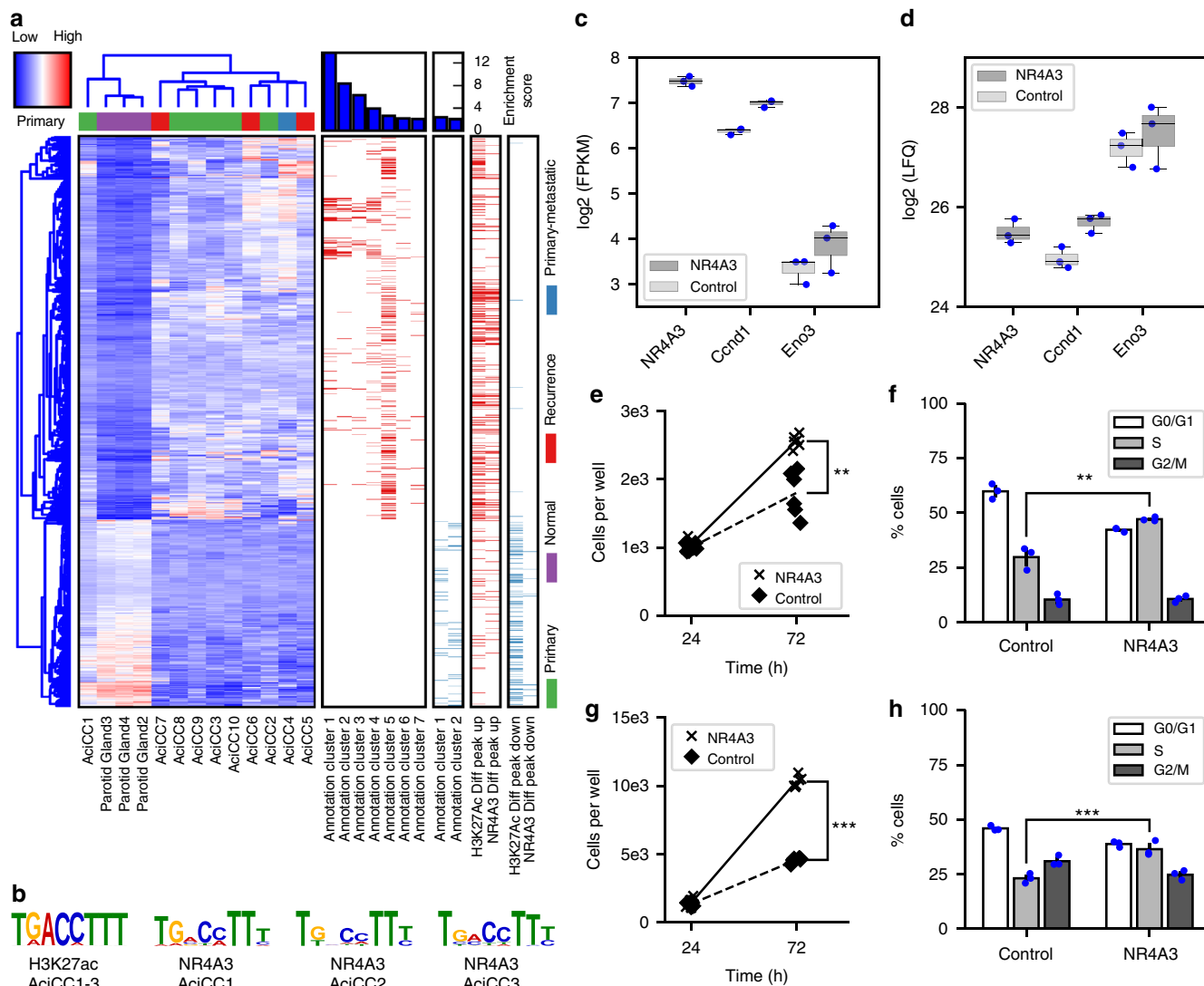
**Fig. 5** Analysis of enhancer activity in a genomic region shared by AcicCs with t(4;9)(q13;q31) rearrangement. Correlation of enhancer activity determined in the dual luciferase reporter assay for ten different genomic fragments (upper panel) with chromatin marks determined by ChIP-seq analyzes in parotid salivary gland tissue (middle panel) and putative enhancer regions based on genome wide chromatin segmentation (lower panel)<sup>(11)</sup>. For each fragment, relative expression values for three replicates normalized to the control vector are presented. Box-plot center line: median; bounds of box: 25 and 75% quantiles; whiskers: minimum and maximum values). One-tailed Student's *t*-test; Amplicon 1:  $P < 0.001$ . Source data are provided as a Source Data file

different in vitro cell models. For the current study, we focused on CCND1 and ENO3 and observed consistently higher mRNA expression and protein levels in human AcicC tissues. Immortalized mouse salivary gland cells stably transduced with human NR4A3 revealed higher mRNA expression and protein levels of mouse Ccnd1 and Eno3 compared to controls, thus confirming a transcriptional effect of NR4A3 upregulation on its known target genes in this cell model. Upregulation of the cell cycle regulator CCND1 is well in line with published data showing that NR4A3 is a potent activator of vascular smooth muscle cell and hepatocyte proliferation<sup>21–23</sup>, with Cyclin D1 representing a direct transcriptional target of NR4A3<sup>13,22,23</sup>. This also correlates with an earlier study reporting on the expression of Cyclin D1 in AcicC<sup>24</sup>. Employing two different cell models derived from mouse salivary gland cells and human mammary gland cells, we observed a stimulatory functional effect of NR4A3 overexpression on cell proliferation and the cell cycle in both models, further supporting its potential oncogenic role. However, we also observed a strong impact of NR4A3 on the expression of genes involved in inflammatory response and complement activation as well as metabolism, referring to the known function of NR4A3 in the transcriptional regulation of inflammatory genes and metabolism<sup>12,14,25</sup>. Higher expression of the glycolytic enzyme ENO3 in human AcicC tissues and our mouse salivary gland cell model with stable NR4A3 overexpression thus supports an additional positive effect of NR4A3 on cell metabolism in AcicC.

NR4As belong to the immediate-early genes, a group of genes (mostly transcription factors) whose expression is rapidly but transiently induced by external signals within minutes to initiate a

rapid but transient cellular response to environmental stimuli<sup>26–28</sup>. Consistently, in a human parotid gland tissue sample we observed both activating and repressive chromatin marks in the promoter region and first intron of NR4A3 indicative of a bivalent/poised promoter. Notably, rapid induction of NR4A3 expression through histone acetylation of upstream regulatory elements has been reported in vascular smooth muscle cells<sup>29</sup>. The high frequency and specificity of the [t(4;9)(q13;q31)] aberration in AcicCs together with the consistent and specific upregulation of NR4A3, but not the neighboring genes, strongly suggests that NR4A3 is indeed the oncogenic driver in AcicCs. A recent study using transcriptomic sequencing analysis reported on a gene fusion on the mRNA level involving the SSCP gene cluster member *Histatin 3* (HTN3) and the gene *Myb/SANT DNA Binding Domain Containing 3* (MSANTD3) in one case of AcicC<sup>9</sup>. In the current study, we could not find any coding chimeric fusion transcripts at the mRNA level, even though one case (AcicC2) displayed an intragenic breakpoint involving the SSCP gene cluster member *Follicular Dendritic Cell Secreted Protein* (FDCSP). This, however, did not result in a chimeric fusion transcript. Instead, we suggest that the highly active chromatin marks we have observed at the SSCP salivary gland gene cluster enable upregulation of NR4A3 through inappropriate regulation of the NR4A3 gene. Indeed, the comparison of 15 AcicCs with the [t(4;9)(q13;q31)] aberration revealed a relatively broad distribution of the 4q13 breakpoints among the SSCP salivary gland gene cluster and no minimal overlapping genomic region was present across all cases, excluding the possibility of a single shared genomic region influencing the expression of





**Fig. 6** Functional impact of NR4A3 upregulation in AciCCs. **a** Hierarchical clustering of ten AciCCs and three normal parotid gland samples based on significantly up- and downregulated genes (deSeq2<sup>47</sup>; FDR = 0.01). Each row represents a gene, and each column represents a sample, with color-coding of gene expression levels. With the exception of one case, the tumors cluster separately from the normal tissues. Functional enrichment analysis reveals seven annotation clusters among upregulated genes, and two annotation clusters among downregulated genes. Genes associated with annotation clusters and differential H3K27ac and NR4A3 peaks are indicated by bars. Red and blue bars encode up- and downregulated genes associated with functional annotation clusters (left), and genes associated with up- and downregulated H3K27ac and NR4A3 peaks (right), respectively. **b** The NBRE motif is the only significantly enriched known motif at sites with upregulated H3K27ac peaks in the tumor samples (HOMER<sup>54</sup>;  $P < 0.01$ ). De-novo motif analysis of NR4A3 transcription factor ChIP-seq peaks in the tumor samples reveals the same NBRE motif in all three samples. **c** mRNA expression and **d** protein levels of NR4A3, Ccnd1, and Eno3 based on RNA sequencing and mass spectrometry analysis, respectively, in immortalized mouse submandibular gland cells transduced with NR4A3 or red firefly luciferase (RedFF) negative control. Box-plots and individual values for three replicates are shown (Box-plot center line: median; bounds of box: 25 and 75 % quantiles; whiskers: minimum and maximum values). **e**, **f** immortalized mouse submandibular gland cells and **g**, **h** human mammary MCF10A cells with and without overexpression of NR4A3 open reading frames were seeded in microscopy plates and incubated for the indicated times. **e**, **g** Cell numbers were counted using high-content screening microscopy after staining with Hoechst-33258. Individual values of six replicates for each time point are presented. **f**, **h** For cell cycle analysis, cells were labeled with BrdU and 7-AAD and analyzed by FACS. Averages and standard deviations as well as individual values of three replicates are shown. Two-tailed Student's *t*-test; \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . Source data for Figs. 1a and 1c–h are provided as a Source Data file

NR4A3 in all samples. In contrast, we identified three distinct subgroups according to the location of the breakpoints within specific genomic regions, and different mechanisms of regulatory influence of 4q13 genomic regions are possible. We focused on the largest group of tumors and functionally validated a distal enhancer located within a 4q13 genomic region included in the translocation in the majority of the cases, thus indirectly presenting reasonable evidence that indeed the 4q13 enhancer activity is the major factor enabling NR4A3 upregulation. In

conclusion we suggest that our comprehensive patient and in vitro data provides strong evidence that the recurrent [t(4;9) (q13;q31)] aberration enables upregulation of NR4A3 through enhancer hijacking and constitutes the initial oncogenic driver event in AciCC. This concept mirrors structural genomic variations in other tumors that have been shown to disrupt spatial organization and activate oncogenes through inappropriate influence of regulatory regions or insulator dysfunction<sup>6–8</sup>. In view of pharmacologic compounds showing antagonistic effects

on NR4A family members<sup>30</sup>, our observations could pave the way for the development of novel treatments through modulation of NR4A3 activity in AcICC in the future.

## Methods

**Patient cohort and informed consent.** Ten patients with diagnosis of AcICC of the salivary glands were enrolled in the initial study cohort, and their tumor material underwent extensive comprehensive genetic and epigenetic profiling. Eight patients gave written informed consent to participate in this study. In the two other patients who had deceased before the study start, archived frozen tumor material was analyzed and written informed consent was given by family members. The study was approved by the local ethics committee of the Friedrich-Alexander University Erlangen-Nuremberg (355\_17 Bc). The remaining tumor samples used for FISH, hybrid capture next generation sequencing and immunohistochemistry were collected from the archives of the Institute of Pathology in accordance with ethical guidelines for the use of retrospective tissue samples by the local ethics committee of the Friedrich-Alexander University Erlangen-Nuremberg (ethics committee statements 24.01.2005 and 18.01.2012).

**Immunohistochemistry.** After formalin-fixation and paraffin-embedding of tumor material, a tissue microarray was conducted with two 2 mm cores per tumor using an automated tissue microarrayer (TMA Grandmaster, 3DHitech, Hungary). Immunohistochemistry was performed on 3 µm sections freshly cut from the TMA block using a fully automated staining system ("Benchmark XT System", Ventana Medical Systems Inc, 1910 Innovation Park Drive, Tucson, Arizona, USA) and monoclonal antibodies against human NR4A3 (1:100 dilution, sc-393902 (H-7), Santa Cruz Biotechnology, INC., Heidelberg, Germany) against human Cyclin D1 (1:50 dilution, clone SP4, Zytomed, Berlin, Germany), and rabbit polyclonal against human Enolase 3 (Beta, Muscle) (1:2000 dilution, HPA000793, Sigma-Aldrich Chemie GmbH, Steinheim, Germany). Cases that were negative for NR4A3 on the TMA slides were subjected to staining of freshly cut whole tissue slides and were re-evaluated. Semi-quantitative scoring for NR4A3 and Enolase 3 immunostaining was evaluated as 0, negative; 1+, weak/focal staining; 2+, intermediate/diffuse staining, and 3+, strong and diffuse staining. The percentage of positively stained nuclei for Cyclin D1 was counted. For NR4A3 and Cyclin D1, only nuclear staining was considered positive, while for Enolase 3, cytoplasmic staining was considered positive. Representative images of tumor and normal parotid gland areas were obtained using an Imager.A2 microscope and AxioCam MRc at ×400 magnification (Carl Zeiss Microscopy GmbH, Jena, Germany) with AxioV40x64 V 4.9.1.0 software.

**Fluorescence in situ hybridization (FISH).** FISH was performed on freshly cut sections from tumor tissue blocks using the ZytoLight SPEC NR4A3 Dual Color Break Apart Probe (ZytoVision GmbH, Bremerhaven, Germany) with standard protocols according to the manufacturer's instructions. Fifty tumor cells were visually inspected using a fluorescence microscope and a cut-off of >20% tumor cells with aberrant signals was used to classify samples as NR4A3-rearranged. All cases were independently scored by two pathologists (F.H., A.A.) and revealed concordant results. The image presented in Supplemental Figure S2 was obtained with Leica SP5 II, Software Revision 2.6.3.1873 using a glycerol objective lens (×63, NA 1.3), detection with PMTs. Nyquist criterion matched for xy and z for all three channels, super critical pinhole with 0.8 Airy Units each channel. Stacks were post processed using Huygens Prof. Vers. 18.04 with deconvolution parameters: CMLE algorithm, SNR:20, 40 iterations.

**Whole genome sequencing.** DNA from AcICC tumors and from normal salivary glands was isolated from fresh-frozen tissues using DNeasy Tissue kits, and from blood of the same patients using the DNeasy Blood Kit (all Qiagen, Hilden, Germany). DNA-quality and quantity were assessed using a 2200 TapeStation (Agilent, Waldbronn, Germany). A total of 1 µg DNA was fragmented (E220; Covaris, Woburn, MA, USA) from each sample and size selected to 300 base pairs (bp) using an E-Gel system (Life Technologies, Darmstadt, Germany). DNA sequencing libraries were prepared using the TruSeq NanoDNA Library PrepKit (Illumina, San Diego, CA, USA) following the manufacturer's instructions. Paired-end sequencing (2 × 150 bp) was performed using one lane of a HiSeqX (Illumina) for every sample.

Reads were aligned against the phase II reference sequence of the 1000 Genomes Project including decoy sequences d5 ([ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2\\_reference\\_assembly\\_sequence/hs37d5.fa.gz](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz)) using bwa mem (version 0.7.8)<sup>31</sup> with default parameters, except for invoking -T 0. After alignment duplicates were marked by applying either Picard (version 1.125) or sambamba (version 0.5.9)<sup>32</sup>. The use of sambamba was necessary, since Picard did not finish for two samples. Genome coverage depth was calculated by custom Perl scripts, described elsewhere<sup>33</sup>. We called SNVs using our in-house, samtools-mpileup based variant calling pipeline<sup>33,34</sup>, that was also used in the ICGC PanCancer<sup>35</sup> project. The workflow can be accessed via the Dockstore webpage: <https://dockstore.org/containers/quay.io/pancancer/pcawg-dkfkz-workflow>.

In brief the workflow first determined variants in the tumor samples and compared these to a matched control to distinguish between somatic and non-somatic calls. The raw calls were further annotated using several publicly available data tracks, such as 1000 Genome variants, single nucleotide polymorphisms database (dbSNP), repeats, and other elements. For annotation of functional impact we used Annovar<sup>36</sup> and further assessed for the variants confidence. Raw calls for indels were obtained from Platypus (v.0.7.4)<sup>37</sup>. Annotation and confidence assessment was done similar to SNV processing. Calculation of genomic translocations was done using the CREST software<sup>38</sup>. In short, we first extracted soft clipped parts of the aligned reads, and then applied crest following the default parameters. In order to filter out germline events we used the -g option allowing for a comparison with a control bam file. In case we had no matched control a fake control was used to filter noise. The resulting translocation events were then annotated using annovar<sup>36</sup>. For visualizing the genomic context of the translocations we defined common TAD regions on the basis of 29 published HiC maps<sup>10</sup>. In summary, we first segmented the genome into 40 kb bins, and counted how many of the 29 samples contained a TAD boundary within each bin and its directly neighboring bins. If the count exceeded a number of 2 we checked if the next bin and its direct neighbors also exceeded a number of two and extended the boundary region by the end of the next bin. This was repeated until a subsequent bin did not fulfill the count criterion anymore, and then the TAD boundary position was defined as the boundary region start position plus the length of the boundary region divided by two. Furthermore, we used the HiC map of the mesoendoderm cell line replicate two for visualizing fine grained TAD structures. Allele-specific copy-number alterations were detected using ACEseq (allele-specific copy-number estimation from WGS)<sup>39</sup>. ACEseq determines absolute allele-specific copy numbers as well as tumor ploidy and tumor cell content based on coverage ratios of tumor and control as well as the B-allele frequency (BAF) of heterozygous SNPs. SVs called by CREST<sup>38</sup> are incorporated to improve genome segmentation.

**Whole genome bisulfite sequencing.** DNA from AcICC tumors were isolated from fresh-frozen tissue using DNeasy Tissue kits and from blood of the same patients using the DNeasy Blood Kit (all Qiagen, Hilden, Germany). DNA-quality and quantity were assessed using a 2200 TapeStation (Agilent, Waldbronn, Germany). A total of 100 ng DNA was fragmented (E220; Covaris, Woburn, MA, USA) from each sample and size selected to 300 bp using an E-Gel system (Life Technologies, Darmstadt, Germany). DNA sequencing libraries were prepared using the Accel-NGS® Methyl-Seq DNA Library Kit (Swift Biosciences, Ann Arbor, MI, USA) following the manufacturer's instructions. Paired-end sequencing (2 × 150 bp) was performed using one lane of a HiSeqX (Illumina) for every sample. Raw paired-end sequence reads were trimmed using Trimmomatic (v. 0.36)<sup>40</sup> with default parameter values to eliminate remnants of the sequencing adapters. Read pairs for which both mates survived the trimming were aligned to the GRCh37 assembly of the human genome reference sequence using bwameth (v. 0.2.0)<sup>41</sup>, a wrapper of the bwa mem<sup>1</sup> alignment algorithm suited for bisulfite sequencing data. Redundant reads emerging during PCR amplification were removed using MarkDuplicates tool of the Picard suite (v. 2.5.0-1) (<http://broadinstitute.github.io/picard>). Methylation states of the CpG cytosines were called with MethylDackel (v. 0.2.1) (<https://github.com/dpryan79/MethylDackel.git>). During methylation calling 10 base pairs were disregarded from 5'-ends of both read mates to eliminate methylation bias artefacts.

**RNA sequencing.** RNA was isolated from AcICC tumors and from cell lines derived from mouse SG (see below) and stably transfected with NR4A3 or red firefly luciferase ORFs, respectively, using the RNeasy Plus Mini Kit (Qiagen). Quality of RNA was analyzed using Agilent Bioanalyser 2100 (Agilent). A volume of 1 µg total RNA was fragmented to a median length of 300 bp (Covaris). Library preparation for RNA-seq was performed using the TruSeq RNA v2 kit for tumor RNAs and the Illumina TruSeq stranded total RNA protocol (Illumina) and NEBNext Multiplex Oligos (New England Biolabs) for cell line RNAs, following the manufacturers' instructions. Sequencing was with paired-end sequencing with 2 × 100 bp HiSeq 4000 and 2 × 125 bp HiSeq2000 v4 technologies (Illumina) for tumor and cell line samples, respectively. Three to four libraries were pooled per lane. Reads were aligned against a STAR reference index generated from the 1000 genomes phase 2 assembly, gencode 19<sup>42</sup> gene models and for a sjdbOverhang of 200. This action was performed by a 2 pass alignment using the STAR aligner (version 2.5.2b)<sup>43</sup>. The alignment call parameters were:

```
--sjdbOverhang 200 --runThreadN 8 --outSAMtype BAM Unsorted
SortedByCoordinate --limitBAMsortRAM 10000000000 --outBAMsortingThreadN
= 1 --outSAMstrandField intronMotif --outSAMunmapped Within KeepPairs
--outFilterMultimapNmax 1 --outFilterMismatchNmax 5
--outFilterMismatchNoverLmax 0.3 --twopassMode Basic --twopass1readsN -1
--genomeLoad NoSharedMemory --chimSegmentMin 15 --chimScoreMin 1
--chimScoreJunctionNonGTAG 0 --chimJunctionOverhangMin 15
--chimSegmentReadGapMax 3 --alignSjstitchMismatchNmax 5 -1 5 5
--alignIntronMax 1100000 --alignMatesGapMax 1100000 --alignSJDBoverhangMin
3 --alignIntronMin 20 --clip3pAdapterSeq AGATCGGAAGAGCACACGTCTGAACTCCAGTCA --readFilesCommand
```

Other parameters were as default. Duplicate marking and index generation on the resulting alignment, and chimeric bam files was performed using sambamba (version 0.6.5)<sup>32</sup>. For sorting we used samtools sort (version 1.3.1)<sup>44</sup>. Quality control analysis was done by applying the samtools flagstat<sup>44</sup> command, and the rnaseqc tool (version 1.1.8)<sup>45</sup> with the 1000 genomes phase 2 assembly and gencode 19 gene models as input. Depth of Coverage analysis for rnaseqc was turned off. Featurecounts (version 1.5.1)<sup>46</sup> was used to perform gene specific read counting over exon features based on the gencode 19 gene models. Both reads of a paired fragment were used in a strand-unspecific manner for counting and the quality threshold was set to 255 (which indicates that STAR found a unique alignment). A custom script was used to calculate FPKM and TPM expression values. For total library abundance calculations, all genes on chromosomes X, Y, MT, as well as rRNA and tRNA genes were omitted as they are likely to introduce library size estimation biases. We analyzed our RNA-sequencing samples for differential expression (tumor vs. normal) by applying the DeSeq2 R package<sup>47</sup> on per sample read counts as determined by the featurecounts software<sup>17</sup>. We removed all genes that had no read counts among our samples from further analysis. Furthermore, we determined the regularized log transformed read count matrix for visualization of the data. Differential analysis was done on 10 tumor samples vs. 3 normal samples. We only considered differentially expressed genes with a FDR smaller or equal to 0.01. Pathway analysis was performed using the DAVID web service<sup>48</sup> on the HUGO gene symbols of differentially expressed genes (FDR <= 0.01). We included the following gene sets into our analysis: From functional categories we used COG\_ONTOLOGY, UP\_KEYWORDS, and UP\_SEQ\_FEATURE. From gene ontology we used GOTERM\_BP\_DIRECT, GOTERM\_CC\_DIRECT, and GOTERM\_MF\_DIRECT, and from pathways we used KEGG\_PATHWAY. For annotation clustering we set the confidence threshold to an EASE score of 0.1, and for the annotation chart we set the EASE score threshold to 0.01. We repeated the analysis independently for up- and downregulated genes with respect to the tumor. For the detection of gene fusions we applied deFuse<sup>49</sup> and the tool arriba (<https://github.com/suhrig/arriba>) using default parameters.

**ChIP sequencing.** Illumina sequencing libraries were prepared from the ChIP and Input DNAs by the standard consecutive enzymatic steps of end-polishing, dA-addition, and adaptor ligation using the Apollo 342 from Wafergen Biosystems. After a final PCR amplification step, the resulting DNA libraries were quantified and sequenced on Illumina's NextSeq 500 (75 nt reads, single end). ChIP-sequencing reads were aligned against the 1000 genomes phase 2 assembly using bwa aln (version 0.7.12)<sup>1</sup> with default parameters. Afterwards duplicate removal was performed by samtools rmdup (version 0.1.19)<sup>44</sup> with parameter -s. Since the 5'-ends of aligned reads (= "tags") represent the end of ChIP/IP-fragments, the tags were extended in-silico (using custom software) at the 3'-ends to a length of 150–250 bp, depending on the average fragment length in the size selected library (normally 200 bp). To identify the density of fragments (extended tags) along the genome, the genome was divided into 32 bp bins and the number of fragments in each bin was determined. This information was stored in bedGraph files and used for visualization. For finding regions of local enrichment in ChIP-seq read density, so-called peaks, we used MACS (version 2.1.0)<sup>50</sup> with parameters effective genome size = 2.7e + 09, band width = 200, model fold = [5, 50], p-value cutoff = 1e-07, and broad calling = off, or SICER (version 1.1)<sup>51</sup> with parameters window size = 200, fragment size = 200, effective genome size as a fraction of hs37d5 = 0.86, gap size = 0, evaluate for identification of candidate islands that exhibit clustering = 1000, FDR = 1e-10, depending on the expected breadth of the peaks. Before applying peak calling, we normalized all samples for a particular ChIP/IP by sampling down all samples to have the same tag counts as the sample with the lowest tag count. Afterwards, we removed peaks residing in ENCODE black-listed regions. Since MACS was particularly made for finding narrow peaks we used it for peak calling on ChIP-Seq data of H3K27ac, H3K4me3, CTCF, and NR4A3. SICER was only used for broad H3K27me3 signals. For the identification of super enhancers we used a proprietary software implemented by Active Motif that gives very similar results as the ROSE super-enhancer software<sup>52,53</sup>. In a first step, MACS peaks of H3K27ac were merged if their distance was equal or less than 12,500 bp. In a second step, the merged peak regions with the strongest signals (top 5%) were identified as super enhancers. For defining differential peaks between tumor and normal samples we first calculated for each histone mark extended regions for which at least one of the samples contains a peak (i.e., the union of peaks). Afterwards, we calculated for each sample and each extended peak interval the average fragment density among all bins within the region. This enabled us to determine the log<sub>2</sub> fold change of average fragment density between tumor samples and the normal sample. We then defined the intervals with log<sub>2</sub> ratio smaller or equal than -2 as downregulated in tumors and the intervals with log<sub>2</sub> ratio greater or equal than 2 as upregulated in tumors. Detection of enriched TFBS motifs in differential peaks was performed using the HOMER software<sup>54</sup>. The perl script findMotifsGenome.pl was applied using default parameters except for -size given and -keepFiles. As background we used the complete set of extended regions per histone mark. We scanned for CTCF motifs within the individual samples' peaks using the motif package of the BioPython module<sup>55</sup>. The position weight matrix (PWM) of CTCF was obtained from the HOCOMOCO database<sup>56</sup>. To avoid overfitting of the PWM we introduced pseudocounts

according to an approximate CG content in the human genome of about 40%. As scanning background we used a global distribution of A,T = 30% and C,G = 20%. Scanning of the motifs according the PWM was done using a precision of 10e4 and a FPR of 10e-4. De-Novo motif search on NR4A3 peak regions was performed using the MEME<sup>57</sup> web interface (<http://meme-suite.org/>). First we extracted for each tumor sample the sequences of the 1000 top ranking peak regions, based on the maximum signal value within the peak. On these the MEME algorithm was applied using the default parameters, except for Minimum Motif With = 8, Maximum Motif Width = 20, Minimum Sites per Motif = 100, and Maximum Sites per Motif = 300.

**Hybrid capture sequencing.** DNA was extracted from freshly cut slides from formalin fixed and paraffin embedded tumor blocks after manual microdissection of tumor areas using the Qiagen QIAamp DNA FFPE Tissue Kit. Libraries were prepared with a customized hybrid capture based approach using probes directed against a ~315 kbp region upstream of the NR4A3 TSS and the SureSelectXT HS Target Enrichment System (Agilent). After quality control, libraries were sequenced on NextSeq 550 (Illumina). Alignment and variant calling was done with the SureCall software for the customized hybrid capture approach (Agilent). Translocations were filtered by a quality score greater or equal to 255, and by a minimal support of the translocation of 15 reads. The remaining translocations were visually inspected using the Integrated Genomics Viewer (IGV) software<sup>58,59</sup> and further filtered.

**Dual luciferase reporter assays.** Genomic regions-of-interest were PCR-amplified from commercial human DNA (Roche, catalog no. 11691112001) using the Q5 kit (New England Biolabs), cloned into reporter vector pGL4.23 (Promega, catalog no. E841A) and sequence verified by Sanger sequencing (GATC Biotech). About 8000 293FT cells were transfected using the TransIT LT1 transfection reagent (Mirus, catalog no. MIR 2304) with 50 ng plasmid mix consisting of 20 ng of the reporter construct or empty vector pGL4.23, 10 ng of the pRL-TK Renilla luciferase reporter vector (Promega, catalog no. E2241) and, as stuffer DNA, 20 ng of a pGL3 (Promega) deletion derivative with a non-functional luciferase gene. The dual luciferase readout was performed with nearly confluent cells using the SpectraMax M5 (Berthold, Bad Wildbad, Germany); the firefly luciferase reporter signals were normalized to the renilla luciferase transfection control signals. Mean values were calculated from six parallel measurements per transfection (technical replicates), and at least three independent transfections were performed per construct (biological replicates). Statistical evaluation of differences between mean values and empty vector (values of normalized signals set to 1) was performed by one-sample *t*-test.

**Overexpression of NR4A3 in mouse SG cells.** Three-week-old female C57BL/6N mice were obtained from Envigo CRS GmbH (Rosdorf, Germany) maintained under standard housing conditions in the DKFZ animal facility for one week. After acclimatization the mice were euthanized by CO<sub>2</sub> inhalation for isolation of SGs. Performance was in accordance with European guideline 2010/63/EU and the guidelines of the local authorities of the DKFZ. SG cells were prepared as described elsewhere<sup>60</sup> with minor modifications. In brief; SGs were excised and transferred to ice-cold phosphate buffered saline (PBS; Thermo Fischer Scientific/ Gibco-BRL, Germany) enriched with 5% fetal calf serum (FCS; Thermo Fischer Scientific/ Gibco-BRL, Germany) and 1% penicillin-streptomycin (Thermo Fischer Scientific/ Gibco-BRL, Germany). Residual blood was removed by two washing steps in ice-cold PBS. SGs were minced in Dulbecco's modified Eagle's medium (DMEM, Thermo Fischer Scientific/ Gibco-BRL, Germany) supplemented with 10% FCS. Minced SG tissue was enzymatically digested for 1 h at 37 °C in 5 mL DMEM supplemented with collagenase type-1 (15 mg; Merck, Germany), hyaluronidase type I-S (5 mg; Merck, Germany), 10% FBS, and DNase I (1 U/mL; Qiagen Germany), and vortexed in 10 min intervals. Cells and tissue fragments were separated by filtering through a 70 cm cell strainer (BD Biosciences). Cells were washed twice in ice-cold PBS (500 g/5 min) and the supernatant of each wash was discarded. The SG cell pellet was suspended in epidermal keratinocyte medium (CnT-07; CELLnTEC Advanced Cell Systems, Bern, Switzerland) containing 0.01 mg/mL human recombinant epidermal growth factor (EGF; Merck, Germany), 2.5 mg/mL penicillin-streptomycin, and 0.1 mg/mL CT (Merck, Germany) in a concentration of 10<sup>6</sup> cells/mL and 10<sup>7</sup> cells in total were seed/10 cm<sup>2</sup> dish coated with 0.1% gelatin (Merck, Germany) and incubated at 37 °C and 5% CO<sub>2</sub>. Medium was changed every two days and the cells were cultured in epidermal keratinocyte medium containing the respective antibiotics. At 80% density the SG cells were passaged 1:4 using Accumax (Thermo Fischer Scientific, Germany) for detachment and reseeded on gelatin coated plates.

To immortalize primary mouse SG cells, the HPV-E6-E7<sup>61</sup> ORF flanked by attL recombination sites was cloned into a pMK-RQ plasmid also carrying a kanamycin selection marker (Thermo Fischer Scientific). The construct was shuttled by Gateway recombination technology (Thermo Fischer Scientific) in a lentiviral expression vector fusing a c-terminal internal ribosomal binding site (IRES), and harboring a neomycin resistance marker. To analyze the effects of NR4A3 overexpression, the (ORF) of human NR4A3 (NM\_173200) flanked by attL recombination sites was synthesized (GeneArt, Germany) and cloned into the pMK-RQ vector (Thermo Fischer Scientific). For normalization, the ORF of red

firefly luciferase (Thermo Fischer Scientific) was amplified using gene specific primers with flanking attB recombination sites. The resulting PCR-product was recombined in pDONR221 (Thermo Fischer Scientific) and sequence validated. The ORFs of NR4A3 and red firefly luciferase were shuttled into a lentiviral expression vector adding a c-terminal IRES sequence coupled to a puromycin resistance gene by Gateway recombination.

For generation of lentiviral particles, 293FT cells (Thermo Fischer Scientific, Germany) were co-transfected with the lentiviral expression constructs and 2nd generation viral packaging plasmids VSV.G (Addgene #14888) and psPAX2 (Addgene #12260). 48 h after transfection, virus containing supernatant was removed and cleared by centrifugation (5 min/500 g). The supernatant was passed through a 0.45- $\mu$ m filter to remove remaining cellular debris. SG cells were immortalized by transduction with HPV-E6-E7<sup>61</sup> lentiviral particles at 70% confluency in the presence of 10  $\mu$ g/mL polybrene (Merck, Germany). 24 h after transduction remaining viral particles were removed and immortalized cells were selected in the presence of 500  $\mu$ g/mL neomycin. For functional analysis immortalized SG cells were transduced with lentiviral particles for overexpression of either NR4A3 or red firefly luciferase control in the presence of polybrene as described above. Stably transduced cells were obtained by selection with 500  $\mu$ g/mL neomycin and 1  $\mu$ g/mL puromycin (Merck, Germany). The cells were tested for mycoplasma contamination on a regular basis.

**Stable transduction of human mammary gland epithelial MCF10A cell lines with NR4A3.** MCF10A cell line was cultured in complete growth medium under standard culture conditions (37 °C; 5% CO<sub>2</sub>). Cells were authenticated upon arrival (ATCC CRL-10317), as well as after completion of the study (Multiplexion, Heidelberg), and tested for mycoplasma infection on a regular basis. Cells were transduced with lentiviral NR4A3 overexpression particles as described above at 70% confluency. Twenty four hours after transduction remaining viral particles were removed and the cells were selected with 1  $\mu$ g/mL puromycin (Merck, Germany). Single clones were picked and tested for expression of NR4A3 by qRT-PCR. Clones showing high and no detectable NR4A3 overexpression, respectively, were selected for functional experiments testing effects of NR4A3 (Supplemental Figure 10, Source Data file).

**Functional analysis of NR4A3 activities on proliferation and cell cycle.** Proliferation of SG or MCF10A cells with or without overexpression of NR4A3 was assessed by cell counting at different time points after seeding. A total of 1500 cells were seeded into each well of 96-well microscopy plates (Greiner, Germany) and grown for 24 h and 72 h at 37 °C with 5% CO<sub>2</sub>. Hoechst-33258 was added to the cells at a final concentration of 1  $\mu$ M and incubated for 30 min before measurement using an ImageXpress Micro high-content screening microscope (Molecular Devices). Images were analyzed with the built-in analysis app “Count Nuclei”. Average and standard deviation of six replicates for each time point were calculated. The two-tailed Student’s *t*-test was used for statistical analyses of the difference in cell numbers after 72 h.

To analyze cell cycle profiles of SG or MCF10A cells with and without overexpression of NR4A3, 1.5 × 10<sup>5</sup> cells/well were seeded into 6-well plates. Twenty four hours after seeding, the cells were starved for 24 h followed by 24 h treatment with full growth medium. Then, the cells were incubated with BrdU for 2 h before fixation with Cytofix/Cytoperm (BD Biosciences, Heidelberg, Germany) according to the manufacturer’s recommendations. BrdU was stained with a FITC-conjugated anti-BrdU antibody (BD Biosciences, Heidelberg, Germany) and DNA was stained with 7-AAD. Samples were analyzed using a FACSCalibur (Becton-Dickenson) and their distributions in the distinct cell cycle phases G0/G1, S and G2/M phases was assessed based on their fluorescence intensity for both dyes as detailed in Supplementary Figure 11. The two-tailed Student’s *t*-test was used for statistical analyses of the differences in the fraction of cells in S-phase.

**Real-time PCR.** Cells were seeded into transparent 6-well plates and RNA was isolated with mRNeasy Mini kit (Qiagen, Hilden, Germany) the following day at 80% confluency. mRNA quantification was performed with TaqMan®-based qRT-PCR and UPL probes (Roche Diagnostics GmbH, Mannheim, Germany). Primer sequences were left-ggaggggggtgaggtgtt and right-gtggcactcttttggctctca for *Actb* using UPL#71, left-tttcttcagagtcatacagtg and right-tgactccagaaggctctca for *Cend1* using UPL#72, left-gaggggggaactgacactc and right-gagtcttccaccggaaga for *Eno3* using UPL#67, left-gccaatagcaccacaaactc and right-agccacatcgctcagacac for GAPDH using UPL#60, and left-ctcaacaccagagatcttgatta and right-gtagaattgttgacatgctcag for NR4A3 using UPL#66. Data were processed with SDS 2.2.2 software (Thermo Fisher Scientific, Waltham, USA) and quantifications were performed with 2- $\Delta\Delta$ CT method. The two-tailed Student’s *t*-test was used for statistical analyses.

**Immunoblotting.** Protein abundance was quantified with western blotting. Cells were seeded into 6-well plates and proteins were isolated the following day at 80% confluency. Primary antibodies used were Anti-Actin (clone C4) (MP Biomedicals GmbH, Eschwege, Germany) for  $\beta$ -Actin (1:10,000 dilution) and SC-718 (Santa Cruz Biotechnology, Inc., Heidelberg, Germany) for cyclin D1 (1:200 dilution). Bands were visualized with secondary IRDye®680 and IRDye®800-conjugated

antibodies and membranes were scanned with Odyssey scanner and analyzed with Odyssey 2.1 (LI-COR, Lincoln, NE, USA).

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All sequencing data supporting the findings of this study have been uploaded to the European Genome-Phenome archive (<http://www.ebi.ac.uk/ega/>), and are available via the accession number EGAS00001002795. All processed data that support the findings of this study are available in ZENODO with the identifier [10.5281/zenodo.1483691]<sup>62</sup>. A reporting summary for this Article is available as a Supplementary Information file. The source data underlying Figs. 1b, c, e, 5, 6a, c–h, and Supplementary Figures 4, 6, 7, 8, 10 are provided as Source Data file. All other data are available from the corresponding authors upon reasonable request.

Received: 8 February 2018 Accepted: 12 December 2018

Published online: 21 January 2019

## References

- Kawasaki, K. & Weiss, K. M. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc. Natl Acad. Sci. USA* **100**, 4060–4065 (2003).
- Vander Poorten, V. et al. Salivary acinic cell carcinoma: reappraisal and update. *Eur. Arch. Otorhinolaryngol.* **273**, 3511–3531 (2016).
- Weinreb, I. Translocation-associated salivary gland tumors: a review and update. *Adv. Anat. Pathol.* **20**, 367–377 (2013).
- Seethala, R. R. & Stenman, G. Update from the 4th edition of the world health organization classification of head and neck tumours: tumors of the salivary gland. *Head. Neck Pathol.* **11**, 55–67 (2017).
- Persson, M. et al. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. *Proc. Natl Acad. Sci. USA* **106**, 18740–18744 (2009).
- Drier, Y. et al. An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. *Nat. Genet.* **48**, 265–272 (2016).
- Northcott, P. A. et al. Enhancer hijacking activates GF11 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).
- Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
- Barasch, N. et al. Recurrent rearrangements of the Myb/SANT-like DNA-binding domain containing 3 gene (MSANTD3) in salivary gland acinic cell carcinoma. *PLoS ONE* **12**, e0171265 (2017).
- Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
- Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Maxwell, M. A. & Muscat, G. E. The NR4A subgroup: immediate early response genes with pleiotropic physiological roles. *Nucl. Recept. Signal.* **4**, e002 (2005).
- Nomiyama, T. et al. Deficiency of the NR4A neuron-derived orphan receptor-1 attenuates neointima formation after vascular injury. *Circulation* **119**, 577–586 (2009).
- Kim, A. Y., Lim, B., Choi, J. & Kim, J. The TFG-TEC oncoprotein induces transcriptional activation of the human  $\beta$ -enolase gene via chromatin modification of the promoter region. *Mol. Carcinog.* **55**, 1411–1423 (2016).
- Kurakula, K., Koenis, D. S., van Tiel, C. M. & de Vries, C. J. NR4A nuclear receptors are orphans but not lonesome. *Biochim. Biophys. Acta* **1843**, 2543–2555 (2014).
- Beard, J. A., Tenga, A. & Chen, T. The interplay of NR4A receptors and the oncogene-tumor suppressor networks in cancer. *Cell Signal.* **27**, 257–266 (2015).
- Labelle, Y. et al. Oncogenic conversion of a novel orphan nuclear receptor by chromosomal translocation. *Hum. Mol. Genet.* **4**, 2219–2226 (1995).
- Mullican, S. E. et al. Abrogation of nuclear receptors Nr4a3 and Nr4a1 leads to development of acute myeloid leukemia. *Nat. Med.* **13**, 730–735 (2007).
- Marti-Pàmies, I., Cañes, L., Alonso, J., Rodríguez, C. & Martínez-González, J. The nuclear receptor NOR-1/NR4A3 regulates the multifunctional glycoprotein vitronectin in human vascular smooth muscle cells. *FASEB J.* **31**, 4588–4599 (2017).
- Filion, C. et al. The EWSR1/NR4A3 fusion protein of extraskeletal myxoid chondrosarcoma activates the PPARG nuclear receptor gene. *J. Pathol.* **217**, 83–93 (2009).
- Martínez-González, J., Rius, J., Castelló, A., Cases-Langhoff, C. & Badimon, L. Neuron-derived orphan receptor-1 (NOR-1) modulates vascular smooth muscle cell proliferation. *Circ. Res.* **92**, 96–103 (2003).

22. Nomiya, T. et al. The NR4A orphan nuclear receptor NOR1 is induced by platelet-derived growth factor and mediates vascular smooth muscle cell proliferation. *J. Biol. Chem.* **281**, 33467–33476 (2006).
23. Vacca, M. et al. Neuron-derived orphan receptor 1 promotes proliferation of quiescent hepatocytes. *Gastroenterology* **144**, 1518–1529 (2013).
24. Liu, T. et al. Abnormal expression of Rb pathway-related proteins in salivary gland acinic cell carcinoma. *Hum. Pathol.* **36**, 962–970 (2005).
25. Zhao, Y. & Bruemmer, D. NR4A orphan nuclear receptors: transcriptional regulators of gene expression in metabolism and vascular biology. *Arterioscler. Thromb. Vasc. Biol.* **30**, 1535–1541 (2010).
26. Tullai, J. W. et al. Immediate-early and delayed primary response genes are distinct in function and genomic architecture. *J. Biol. Chem.* **282**, 23981–23995 (2007).
27. Pearen, M. A. & Muscat, G. E. Minireview: Nuclear hormone receptor 4A signaling: implications for metabolic disease. *Mol. Endocrinol.* **24**, 1891–1903 (2010).
28. Bahrami, S. & Drabløs, F. Gene regulation in the immediate-early response process. *Adv. Biol. Regul.* **62**, 37–49 (2016).
29. Zhao, Y. et al. Epigenetic regulation of the NR4A orphan nuclear receptor NOR1 by histone acetylation. *FEBS Lett.* **588**, 4825–48230 (2014).
30. Safe, S., Jin, U. H., Hedrick, E., Reeder, A. & Lee, S. O. Minireview: role of orphan nuclear receptors in cancer and potential as drug targets. *Mol. Endocrinol.* **28**, 157–172 (2014).
31. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
32. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
33. Jones, D. T. W. et al. Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).
34. Jones, D. T. W. et al. Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nat. Genet.* **45**, 927–932 (2013).
35. Stein, L. D., Knoppers, B. M., Campbell, P., Getz, G. & Korbel, J. O. Data analysis: create a cloud commons. *Nature* **523**, 149–151 (2015).
36. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
37. Rimmer, A. et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).
38. Wang, J. et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods* **8**, 652–654 (2011).
39. Kleinheinz, K. et al. ACESeq - allele specific copy number estimation from whole genome sequencing. Preprint at [bioRxiv](https://doi.org/10.1101/210807) <https://doi.org/10.1101/210807> (2017).
40. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
41. Pedersen, B. S., Eyring, K., De, S., Yang, I. V. & Schwartz, D. A. Fast and accurate alignment of long bisulfite-seq reads. Preprint at <https://arxiv.org/abs/1401.1129> (2014).
42. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
43. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma. (Oxf. Engl.)* **29**, 15–21 (2013).
44. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinforma. (Oxf. Engl.)* **25**, 2078–2079 (2009).
45. DeLuca, D. S. et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinforma. (Oxf. Engl.)* **28**, 1530–1532 (2012).
46. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
47. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
48. Huang, D. W. et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* **8**, R183 (2007).
49. McPherson, A. et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-seq data. *PLoS Comput. Biol.* **7**, e1001138 (2011).
50. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
51. Zang, C. et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).
52. Lovn, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
53. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
54. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
55. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
56. Kulakovskiy, I. V. et al. HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* **41**, D195–D202 (2013).
57. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
58. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
59. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. O. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
60. Ikeura, K., Kawakita, T., Tsunoda, K., Nakagawa, T. & Tsubota, K. Characterization of long-term cultured murine submandibular gland epithelial cells. *J. Virol.* **65**, 473–478 (1991).
61. Halbert, C. L., Demers, G. W. & Galloway, D. A. The E7 gene of human papillomavirus type 16 is sufficient for immortalization of human epithelial cells. *J. Virol.* **65**, 473–478 (1991).
62. Haller, F. et al. Enhancer hijacking activates oncogenic transcription factor NR4A3 in Acinic Cell Carcinomas of the salivary glands (datasets). ZENODO. *Broadminded Blog.* <https://doi.org/10.5281/zenodo.1483691> (2018).

## Acknowledgements

We thank Bishaka Roy and Andrea Waxmann from the sequencing and Bernd Hessling and Martin Schneider from the protein analysis units of the DKFZ Genomics and Proteomics Core Facility for performing excellent services. We thank Christa Winkelmann and Susanne Blank from the Institute of Pathology for excellent contribution to immunohistochemical and in situ analyses and Heike Wilhelm for excellent technical assistance with molecular and cellular biology experiments. We thank the DKFZ-Heidelberg Center for Personalized Oncology (DKFZ-HIPO) and Dr. Carl Hermann, Department of Theoretical Bioinformatics, for helpful expertise. Fluorescence microscopy/image analysis was performed with the support of the Optical Imaging Centre Erlangen (OICE) and the excellent expertise of Dr. Ralf Palmisano. Bioinformatic analysis of hybrid capture based sequencing of NR4A3 upstream region was performed with the helpful support of Dr. Michael Walter, Agilent Technologies Sales and Services GmbH.

## Author contributions

M.Bi., N.I., P.L., K.K. and R.E. performed or supervised sequencing data analyses. R.W. and B.K. established cell line models. C.K., A.B., A.W. and S.W. performed in vitro assays and performed data analysis. D.W. and M.Ba. performed and analyzed dual luciferase reporter assay. S.K.M., C.S., M.H., R.F., H.I. and A.H. collected specimens. F.H., E.A.M. and A.A. performed immunohistochemical and in situ analyses. F.H., M.Bi. and E.A.M. performed and analyzed hybrid capture next generation sequencing analyses. F.H., M.Bi., R.W., C.P., S.W. and A.A. interpreted the results. F.H., M.Bi., R.W. and A.B. generated figures and tables. F.H., M.Bi., C.P., S.W. and A.A. conceived the study and wrote the manuscript. All authors participated in discussions and interpretation of the data and results.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-08069-x>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npj.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019