



**HAL**  
open science

## **Chimeric oncogene regulates the EGR2 sarcoma susceptibility gene via a GGAA-microsatellite**

Thomas G P Grünewald, Virginie Bernard, Pascale Gilardi-Hebenstreit, Virginie Raynal, Didier Surdez, Marie-Ming Aynaoud, Florencia Cidre-Aranaz, Olivier Mirabeau, Franck Tirode, Gaëlle Pérot, et al.

### ► To cite this version:

Thomas G P Grünewald, Virginie Bernard, Pascale Gilardi-Hebenstreit, Virginie Raynal, Didier Surdez, et al.. Chimeric oncogene regulates the EGR2 sarcoma susceptibility gene via a GGAA-microsatellite. *Nature Genetics*, 2015, 47 (9), pp.1073-1078. 10.1038/ng.3363 . inserm-02440399

**HAL Id: inserm-02440399**

**<https://inserm.hal.science/inserm-02440399>**

Submitted on 15 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Chimeric oncogene regulates the *EGR2* sarcoma susceptibility gene via a GGAA-microsatellite

Thomas G. P. Grunewald<sup>1</sup>, Virginie Bernard<sup>2</sup>, Pascale Gilardi-Hebenstreit<sup>3</sup>, Virginie Raynal<sup>1</sup>, Didier Surdez<sup>1</sup>, Marie-Ming Aynaud<sup>1</sup>, Florencia Cidre-Aranaz<sup>4</sup>, Olivier Mirabeau<sup>1</sup>, Franck Tirode<sup>1</sup>, Gaëlle Perot<sup>5</sup>, Anneliene H. Jonker<sup>1</sup>, Carlo Lucchesi<sup>1</sup>, Marie-Cécile Le Deley<sup>6</sup>, Odile Oberlin<sup>7</sup>, Perrine Marec-Bérard<sup>8</sup>, Amélie S. Véron<sup>9</sup>, Stephanie Reynaud<sup>10</sup>, Eve Lapouble<sup>10</sup>, Valentina Boeva<sup>11</sup>, Thomas Rio Frio<sup>2</sup>, Javier Alonso<sup>4</sup>, Gaëlle Pierron<sup>10</sup>, Geraldine Cancel-Tassin<sup>12</sup>, Olivier Cussenot<sup>12</sup>, David. G. Cox<sup>9</sup>, Stephen J. Chanock<sup>13</sup>, Patrick Charnay<sup>3</sup> and Olivier Delattre<sup>1,2,10,§</sup>

1 INSERM U830 “Genetics and Biology of Cancers”, Institut Curie Research Centre, 26 rue d’Ulm, 75248 Paris, cedex 5, France

2 Institut Curie Genomics of Excellence (ICGex) platform, Institut Curie Research Centre, 26 rue d’Ulm, 75005 Paris, cedex 5, France

3 Ecole Normale Supérieure (ENS), Institut de Biologie de l’ENS (IBENS), INSERM U1024, CNRS UMR 8197, 46 rue d’Ulm, 75005 Paris, cedex 5, France

4 Instituto de Investigación de Enfermedades Raras, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo Km 2, 28220 Majadahonda, Madrid, Spain

5 INSERM U916 “Biology of Sarcomas”, Institut Bergonié, 229 cours de l’Argonne, 33076 Bordeaux cedex, France

6 Département d’Epidémiologie et de Biostatistiques, Institut Gustave Roussy, 114 rue Édouard-Vaillant, 94805 Villejuif cedex, France

7 Département de Pédiatrie, Institut Gustave Roussy, 114, rue Édouard-Vaillant, 94805 Villejuif cedex, France

8 Institute for Paediatric Haematology and Oncology, Leon-Bérard Cancer Centre, University of Lyon, 1 Place Joseph Renault, 69008 Lyon, France

9 INSERM U1052, Léon-Bérard Cancer Centre, Cancer Research Centre of Lyon, 28 rue Laennec, 69373 Lyon, cedex 8, France

10 Unité Génétique Somatique (UGS), Institut Curie, Centre Hospitalier, 26 rue d’Ulm, 75005 Paris, cedex 5, France

11 INSERM U900 “Bioinformatics and Computational Systems Biology of Cancer”, Institut Curie Research Centre, 26 rue d’Ulm, 75248 Paris, cedex 5, France

12 CeRePP - Laboratory for Urology, Research Team 2, UPMC, Hôpital Tenon, 4 rue de la Chine, 75970 Paris, cedex 20, France

13 Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI), Bethesda, 9609 Medical Center Drive, MSC 9776, Bethesda, Maryland 20892, USA

### § address for correspondence:

Olivier Delattre, M.D., Ph.D.

INSERM U830 “Genetics and Biology of Cancers”

Institut Curie Research Centre

26 rue d’Ulm, 75248 Paris cedex 05, France

Tel.: ++33 (0) 1 56 24 66 79

Fax.: ++33 (0) 1 56 24 66 30

Email: olivier.delattre @curie.fr

**Word count:** 1496 (excluding introductory paragraph, references, and figure legends)

**Display items:** 4 figures, 5 supplementary figures, 9 supplementary tables

**Running title:** Role of *EGR2* in susceptibility to Ewing sarcoma

**Key words:** *EWSR1-FLII*, Ewing sarcoma, germline susceptibility, GGAA-microsatellite, *EGR2*

### **Introductory paragraph**

Deciphering how somatic driver mutations and germline susceptibility variants cooperate to promote cancer constitutes a considerable challenge. Ewing sarcoma (EwS) is hallmarked by chimeric *EWSR1-ETS* gene fusions, such as *EWSR1-FLI1*, encoding potent oncogenic transcription factors that can bind DNA at GGAA-motifs<sup>1-3</sup>. EwS incidence is highly variable across human populations, which was recently explored by a genome-wide association study (GWAS) that highlighted susceptibility variants near *EGR2*<sup>4</sup>. Here we show that *EGR2* is crucial for tumourigenesis of EwS and further identify a variant, which contributes to *EWSR1-FLI1*-driven *EGR2* overexpression. We found that *EGR2* knockdown inhibits proliferation, clonogenicity, and spheroidal growth of EwS cells *in vitro*, and induces complete regression of EwS xenografts *in vivo*. Targeted deep-sequencing of the *EGR2* locus in constitutional DNA of 348 cases and 252 PCA-matched controls revealed 267 single nucleotide polymorphisms (SNPs) significantly associated with EwS. At rs79965208, the A-risk-allele is significantly more frequent in EwS than in controls and is in strong linkage disequilibrium with the nearby most significant EwS-associated SNPs. It connects two adjacent GGAA-repeats by conversion of an interspaced GGATT- into a GGAAA-motif. Thereby, the A-allele considerably increases the number of consecutive GGAA-repeats and thus *EWSR1-FLI1*-dependant enhancer activity of this DNA sequence that exhibits chromatin characteristics of an active regulatory element. Accordingly, the A-allele is associated with increased overall and allele-specific *EGR2* expression in EwS tumours. Collectively, our data establish cooperation between a dominant oncogene and a susceptibility variant that regulates a major driver of EwS tumourigenesis.

**Letter (words 1496, max. 1500)**

EwS is an aggressive paediatric malignancy that likely arises from neural crest- or mesoderm-derived mesenchymal stem cells (MSCs)<sup>5,6</sup>. EwS is driven by oncogenic *EWSR1-ETS* fusions (mostly by *EWSR1-FLI1*)<sup>1,7,8</sup>. *EWSR1-FLI1* binds DNA either at ETS-like consensus sites containing a GGAA core-motif or, more specifically with respect to other ETS family members, at GGAA-microsatellites, where the enhancer activity increases with the number of consecutive GGAA-motifs<sup>2,3</sup>. Notably, ~40% of genomic *EWSR1-FLI1* occupancy appears at GGAA-microsatellites<sup>9</sup>. Besides *EWSR1-FLI1*, few additional somatic abnormalities occur in EwS<sup>10-12</sup>. Several epidemiological studies, however, have documented striking disparities in the incidence of EwS across human populations<sup>13</sup>, implying a strong contribution of germline variation to EwS tumorigenesis. In agreement with this hypothesis, our recent GWAS has identified three very significant susceptibility loci with high odds ratios (OR>2)<sup>4</sup>. However, the potential oncogenic cooperation between the major *EWSR1-FLI1* somatic alteration and these EwS susceptibility loci remained unclear.

To further investigate this aspect, we focused on the chr10q21.2-3 susceptibility locus, containing *ADO* (*2-aminoethanethiol dioxygenase*) that encodes a non-heme iron enzyme converting cysteamine into taurine<sup>14</sup>, and *EGR2* (*early growth response 2*, alias *KROX20*), which is a conserved zinc-finger transcription factor promoting proliferation, differentiation, and/or survival in different cell types, including neural crest-derived Schwann cells and mesoderm-derived osteoprogenitors<sup>15-18</sup>. Previous data showed that *ADO* and *EGR2* are overexpressed in EwS as compared to other solid tumours and that elevated expression is associated with risk-alleles<sup>4</sup>. *EGR2*, and to a lesser extent *ADO*, are also overexpressed in EwS as compared to normal tissues ([Fig. 1a](#), [Supplementary Fig. 1](#)). Between-group analysis of microarray data of seven paediatric soft-tissue and brain tumour types demonstrated that *EGR2*, but not *ADO*, clusters with established *EWSR1-FLI1*-target genes<sup>19</sup> ([Fig. 1b](#)). To

further explore the eQTL properties of the chr10 locus, we took advantage of available genotype and matched expression datasets from EwS and other small-round-cell tumours as well as normal tissues. Interestingly, the EwS risk-associated rs1848797, which was genotyped in all datasets, is associated with higher *EGR2* and *ADO* expression only in EwS but not in any *EWSR1-FLI1*-negative tissues (Fig. 1c, [Supplementary Table 1](#)). Moreover, ectopic *EWSR1-FLI1* expression in human MSCs specifically induces *EGR2* (Fig. 1d), while *EWSR1-FLI1* knockdown using specific small interfering RNAs (siRNAs) consistently reduces its expression in four different EwS cell lines ([Supplementary Fig. 2](#)). Such regulation by *EWSR1-FLI1* was not observed for *ADO*. These data strongly suggest that *EGR2* and *ADO* are specifically regulated by eQTL in EwS, but that only *EGR2* is *EWSR1-FLI1*-dependant.

We next asked whether these genes are relevant for tumourigenesis of EwS. Knockdown experiments revealed that inhibition of *EGR2*, but not of *ADO*, impairs proliferation and clonogenicity of four different EwS cell lines, which is paralleled by reduced cell cycle progression through S-phase and viability (Fig. 2a,b, [Supplementary Fig. 3](#)). To confirm the relevance of *EGR2* for EwS growth, we generated EwS cell lines with a doxycycline-inducible anti-*EGR2* small-hairpin RNA (shRNA) expression system. Long-term *EGR2* knockdown not only dramatically reduces anchorage-independent spheroidal growth *in vitro* but, even more strikingly, induces complete regression of EwS xenografts *in vivo* (Fig. 2c,d). In line with *EGR2* being downstream of *EWSR1-FLI1*, transcriptome profiling of EwS cells by DNA microarrays 48 h after knockdown of *EGR2* demonstrated that both genes highly significantly overlap in their transcriptional signatures (Fig. 2e, [Supplementary Tables 2-4](#)). Collectively, these data suggest that *EGR2* is an *EWSR1-FLI1*-induced target gene, which is essential for EwS tumourigenicity.

To fine-map the chr10 susceptibility locus and to identify variants that may contribute to *EGR2* overexpression, we performed targeted parallel deep-sequencing (150/150nt) of the

entire chr10 susceptibility locus and flanking haplotype blocks from constitutional DNA of 348 EwS cases and 252 PCA-matched controls (median target-region coverage  $\geq 10X$ : 92.74%; median nucleotide-coverage: 217X; [Supplementary Fig. 4](#)). Taking into account thresholds of coverage ( $\geq 10X/\text{position}$ ), genotyping rate at observed SNPs ( $\geq 90\%$  of cases) and heterozygosity ( $\text{MAF} \geq 0.05$ ), as well as compliance with Hardy-Weinberg equilibrium (HWE), a total of 266 SNPs located in the previously defined haplotype block demonstrated significant differences in genotype distribution between EwS cases and controls ([Fig. 3a](#), [Supplementary Table 5](#), [Supplementary Fig. 5](#)). In particular, significant associations with EwS were confirmed for all 13 previously identified GWAS SNPs that displayed sufficient coverage<sup>4</sup>. Haplotype and linkage disequilibrium (LD) analysis revealed that this locus may be subdivided into discrete sub-haploblocks ([Fig. 3b](#), [Supplementary Table 5](#)).

To prioritize SNPs for functional assessment, we first informed our sequencing data with published ChIP-seq, DNase-seq, and ENCODE data from EwS cell lines<sup>9,20,21</sup> as recent evidence suggests that most causal SNPs may cluster in epigenetically active and cell type specific regulatory elements<sup>22,23</sup> ([Fig. 3c](#)). Secondly, we included knowledge on evolutionary conserved *EGR2* regulatory elements previously mapped in model animals<sup>24,25</sup>. Indeed, activating chromatin marks, signals for formaldehyde-assisted isolation of regulatory elements (FAIRE), and/or DNase1 hypersensitivity were observed at five main loci: Two loci correspond to known *EGR2* regulatory elements (MSE, Myelinating Schwann cell Enhancer; BoneE, Bone Enhancer), one to the *ADO* promoter, and two to GGAA-microsatellites (mSat1 and mSat2) that overlay with strong EWSR1-FLI1 ChIP-seq signals ([Fig. 3c](#)). As the *ADO* promoter does not contain any EwS-associated SNPs, it was not further investigated. Luciferase reporter assays indicated that BoneE and MSE have no or weak activity in EwS, respectively ([Fig. 3d,e](#)). In contrast, mSat1 and mSat2 both exhibited strong EWSR1-FLI1-dependant enhancer-like activity ([Fig. 3d,e](#)). This activity corresponds to the strictly EWSR1-

FLI1-dependant presence of the activating H3K4me1 and H3K27ac chromatin marks at mSat1 and mSat2 (Fig. 3c), and is consistent with recent data suggesting that EWSR1-FLI1 can act as a pioneer transcription factor creating *de novo* enhancers at GGAA-microsatellites

Due to its higher enhancer activity and its relatively simpler structure as compared to mSat1, as well as to its localization in the sub-haploblock containing the most significant EwS-associated SNPs, we focused on mSat2 and performed PCR-based targeted long-read (300/300nt) deep re-sequencing of mSat2 to analyse in-depth its genetic architecture. This approach yielded n=1178 analysable mSat2 sequences and led to the identification of another SNP being significantly associated with EwS (Fig. 4a, Supplementary Table 6). This SNP, reported as rs79965208, is in strong LD ( $D'=0.97$ ) with the nearby rs6479860, which is the most significant EwS-associated SNP across the whole chr10 susceptibility locus (Fig. 4a, Supplementary Table 5). Interestingly, this SNP converts a GGAT- into a GGAA-motif, thereby connecting two adjacent GGAA-repeats (Fig. 4a). The first GGAA-repeat is polymorphic and contains a median number of eleven GGAA-motifs (range: 7 to 17), while the second is not polymorphic and composed of four GGAA-motifs. The A-allele therefore increases the median number of consecutive GGAA-motifs from eleven to 16. In agreement, considering a previously described threshold for exponentially increasing EWSR1-FLI1-dependant enhancer activity (above twelve consecutive GGAA-motifs)<sup>4</sup>, a significantly larger proportion of EwS mSat2 sequences contains more than twelve GGAA-motifs as compared to controls (65.2 vs. 56.6%,  $P=0.0015$ ). We subsequently tested the enhancer properties of mSat2 corresponding to the reference sequence (hg19) containing either the T- or A-allele at rs79965208 (that is, either containing eleven or 16 consecutive GGAA-motifs) in a luciferase assay. Strikingly, as compared to the T-allele, the A-allele strongly increases the EWSR1-FLI1-induced enhancer activity of mSat2 (Fig. 4b). This transcription activation property is

observed in two different EwS cell lines and strictly depends on the presence of EWSR1-FLI1, since doxycycline-induced knockdown of *EWSR1-FLI1* abrogates or considerably decreases luciferase activity. We next investigated whether the genotype at this SNP is associated with *EGR2* expression levels in EwS tumours. In accordance with the reporter assays, the A-allele is associated with a significantly higher *EGR2* expression in EwS tumours (Fig. 4c). Moreover, taking advantage of a transcribed SNP in the 3'UTR of *EGR2* (rs61865883), we assessed allele-specific expression (ASE) depending on the genotype at rs79965208. Across 45 cases for which ASE could be determined, a significantly higher ASE-rate was observed in cases heterozygous at rs79965208 as compared to homozygous cases (Fig. 4d). Collectively, our results show that *EGR2* is a pro-tumourigenic EwS susceptibility gene whose overexpression in tumours is mediated by EWSR1-FLI1 through a risk-conferring enhancer-like polymorphic GGAA-microsatellite (Fig. 4e).

To our knowledge this constitutes one of the first reports of a pathogenic and functional interplay between a germline polymorphism in strong LD with SNPs responsible for a GWAS signal and a cancer-specific acquired genetic abnormality. Our finding converges with recent predictions that causal variants likely do not range among the most significant ones that led to the identification of the susceptibility locus, but rather are in strong linkage with them and concern transcription regulatory elements<sup>26</sup>. Notably, as determined by the 1000 Genomes project<sup>27</sup>, the A-allele at rs79965208 is far less frequent in African populations (Supplementary Table 7), suggesting that it may contribute to the higher EwS-incidence in individuals of Caucasian and Asian ancestry as compared to those of African ancestry<sup>13</sup>. However, as many additional SNPs in the different sub-haploblocks of the chromosome 10 locus are associated with EwS, it remains likely that other SNPs, in conjunction with rs79965208, also impact *EGR2* expression.

## **Acknowledgements**

T.G.P.G. is supported by a grant from the German Research Foundation (DFG GR3728/2-1). D.S. is supported by the Institut Curie–SIRIC (Site de Recherche Intégrée en Cancérologie) program. D.G.C. is supported by a grant from the InfoSarcomes association. Additionally, this work was supported by grants from the Institut Curie, the INSERM, the ANR-10-EQPX-03 from the Agence Nationale de la Recherche (investissements d'avenir), ANR10-INBS-09-08 from the Canceropôle Ile-de-France, the Ligue Nationale Contre le Cancer (Equipe labellisée), the Institut National du Cancer (PLBIO14-237), the European PROVABES (ERA-649 NET TRANSCAN JTC-2011), ASSET (FP7-HEALTH-2010-259348), and EEC (HEALTH-F2-2013-602856) projects, the Société Française des Cancers de l'Enfant and the following associations: Courir pour Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon, Enfants et Santé, and les Amis de Claire. We thank Dr. John Maris for providing genotype information for the neuroblastoma dataset as well as Dr. Liming Liang and Dr. William Cookson for providing access to genotype data of the LCL dataset. The authors also thank the following clinicians and pathologists for providing samples used in this work: I. Aerts, P. Anract, C. Bergeron, L. Boccon-Gibod, F. Boman, F. Bourdeaut, C. Bouvier, R. Bouvier, L. Brugières, E. Cassagnau, J. Champigneulle, C. Cordonnier, J. M. Coindre, N. Corradini, A. Coulomb-Lhermine, A. De Muret, G. De Pinieux, A.S. Defachelles, A. Deville, F. Dijoud, F. Doz, C. Dufour, K. Fernandez, N. Gaspard, L. Galmiche-Rolland, C. Glorion, A. Gomez-Brouchet, J.M. Guinebretière, H. Jouan, C. Jeanne-Pasquier, B. Kantelip, F. Labrousse, V. Laithier, F. Larousserie, G. Leverger, C. Linassier, P. Mary, G. Margueritte, E. Mascard, A. Moreau, J. Michon, C. Michot, F. Millot, Y. Musizzano, M. Munzer, B. Narciso, O. Oberlin, D. Orbach, H. Pacquement, Y. Perel, B. Petit, M. Peuchmaur, J.Y. Pierga, C. Piguet, S. Piperno-Neumann, E. Plouvier, D. Ranchere-Vince, J. Rivel, C. Rouleau, H. Rubie, H. Sartelet, G. Schleiermacher, C. Schmitt, N. Sirvent, D. Sommelet, P. Terrier, R.

Tichit, J. Vannier, J. M. Vignaud, and V. Verkarre. We also thank David Darmon for excellent technical assistance, and Dr. Veit R. Buchholz and Dr. Elke Butt for critical reading of the manuscript.

### **Author contributions**

T.G.P.G. coordinated and designed the study, performed all functional experiments, analysed the sequencing data, wrote the paper and designed the figures, and helped in grant applications. V. Bernard processed the sequencing data and performed bioinformatic analyses. P.G.-H. participated in study design, performed cloning of risk-alleles and enhancer elements, and contributed to data analysis. V.R. performed all sequencing experiments and helped in analysing the bioinformatic data. D.S. contributed to the *in vivo* experiments and provided experimental protocols. M.M.A., F.C.-A., and A.H.J. helped in functional experiments. O.M., F.T., and C.L. provided statistical advice and helped in the bioinformatic analyses. G. Perot assisted in generation of the shRNA-constructs. M.C.L.D., O.O. and P.M.-B. provided EwS samples and annotation. G. Pierron, S. R., and E.L. provided and prepared EwS samples. T.R.-F. coordinated and supervised sequencing experiments. V. Boeva helped with analysis of the ChIP-seq data. J.A. provided the A673/TR/shEF1 and SK-N-MC/TR/shEF1 cell lines. A.S.V. performed the PCA-clustering of cases and controls. G.C.-T. and O.C. provided DNA of healthy controls. D.G.C. and S.J.C. provided genetic and statistical guidance. P.C. provided advice on analyses concerning *EGR2* and laboratory infrastructure. O.D. initiated, designed, and supervised the study, provided biological and genetic guidance, analysed the data, wrote the paper together with T.G.P.G., and provided laboratory infrastructure and financial support. All authors read and approved the final manuscript.

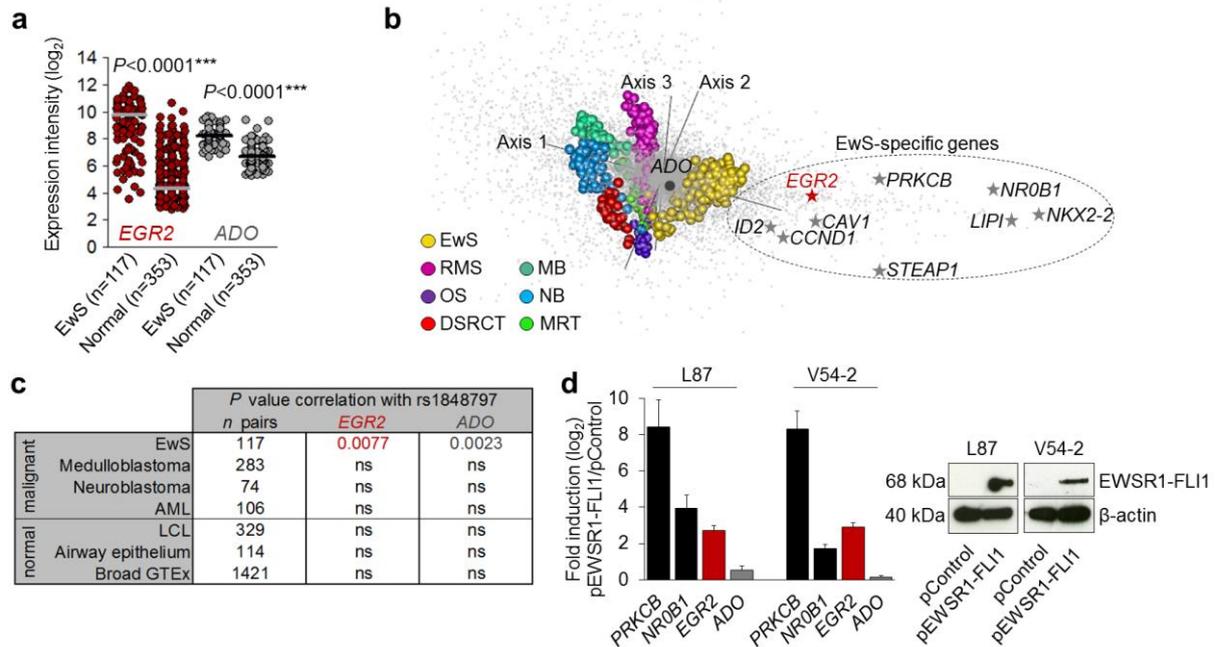
**Competing financial interests**

The authors declare no competing financial interests.

## References

1. Delattre, O. *et al.* Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359**, 162–165 (1992).
2. Gangwal, K. *et al.* Microsatellites as EWS/FLI response elements in Ewing's sarcoma. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10149–10154 (2008).
3. Guillon, N. *et al.* The oncogenic EWS-FLI1 protein binds in vivo GGAA microsatellite sequences with potential transcriptional activation function. *PLoS One* **4**, e4932 (2009).
4. Postel-Vinay, S. *et al.* Common variants near TARDBP and EGR2 are associated with susceptibility to Ewing sarcoma. *Nat. Genet.* **44**, 323–327 (2012).
5. Von Levetzow, C. *et al.* Modeling initiation of Ewing sarcoma in human neural crest cells. *PLoS One* **6**, e19305 (2011).
6. Tirode, F. *et al.* Mesenchymal stem cell features of Ewing tumors. *Cancer Cell* **11**, 421–429 (2007).
7. Delattre, O. *et al.* The Ewing family of tumors--a subgroup of small-round-cell tumors defined by specific chimeric transcripts. *N. Engl. J. Med.* **331**, 294–299 (1994).
8. Sorensen, P. H. *et al.* A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor, ERG. *Nat Genet* **6**, 146–51. (1994).
9. Patel, M. *et al.* Tumor-specific retargeting of an oncogenic transcription factor chimera results in dysregulation of chromatin and transcription. *Genome Res.* **22**, 259–270 (2012).
10. Brohl, A. S. *et al.* The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet.* **10**, e1004475 (2014).
11. Crompton, B. D. *et al.* The Genomic Landscape of Pediatric Ewing Sarcoma. *Cancer Discov.* (2014). doi:10.1158/2159-8290.CD-13-1037
12. Tirode, F. *et al.* Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov.* (2014). doi:10.1158/2159-8290.CD-14-0622
13. Worch, J. *et al.* Racial differences in the incidence of mesenchymal tumors associated with EWSR1 translocation. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **20**, 449–453 (2011).
14. Dominy, J. E., Jr *et al.* Discovery and characterization of a second mammalian thiol dioxygenase, cysteamine dioxygenase. *J. Biol. Chem.* **282**, 25189–25198 (2007).
15. Chandra, A., Lan, S., Zhu, J., Siclari, V. A. & Qin, L. Epidermal growth factor receptor (EGFR) signaling promotes proliferation and survival in osteoprogenitors by increasing early growth response 2 (EGR2) expression. *J. Biol. Chem.* **288**, 20488–20498 (2013).
16. Maro, G. S. *et al.* Neural crest boundary cap cells constitute a source of neuronal and glial cells of the PNS. *Nat. Neurosci.* **7**, 930–938 (2004).
17. Schneider-Maunoury, S. *et al.* Disruption of Krox-20 results in alteration of rhombomeres 3 and 5 in the developing hindbrain. *Cell* **75**, 1199–1214 (1993).
18. Topilko, P. *et al.* Krox-20 controls myelination in the peripheral nervous system. *Nature* **371**, 796–799 (1994).
19. Mackintosh, C., Madoz-Gúrpide, J., Ordóñez, J. L., Osuna, D. & Herrero-Martín, D. The molecular pathogenesis of Ewing's sarcoma. *Cancer Biol. Ther.* **9**, 655–667 (2010).

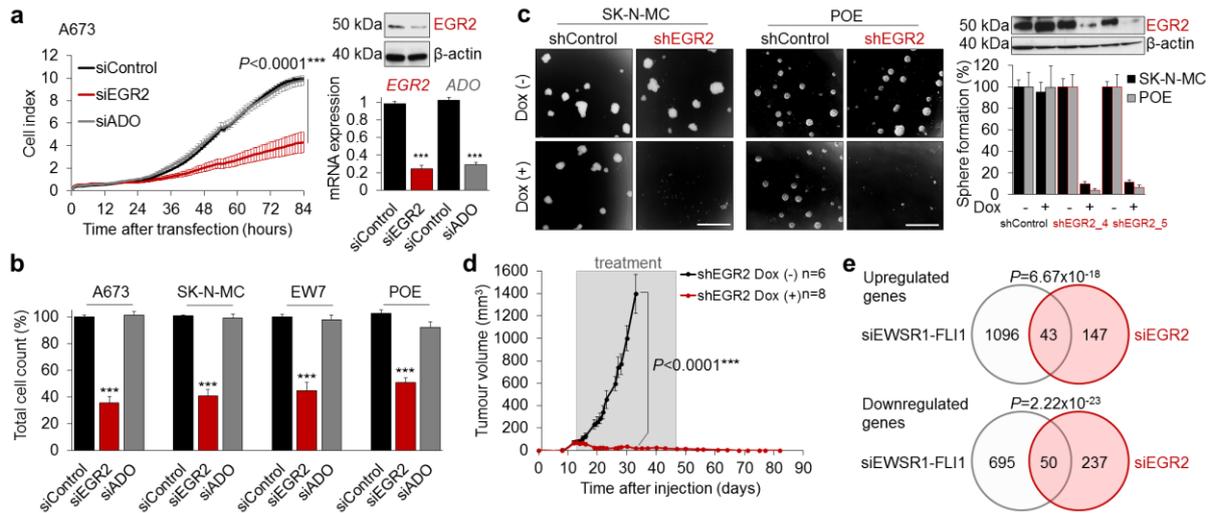
20. Riggi, N. *et al.* EWS-FLI1 Utilizes Divergent Chromatin Remodeling Mechanisms to Directly Activate or Repress Enhancer Elements in Ewing Sarcoma. *Cancer Cell* **0**,
21. ENCODE Project Consortium *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
22. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
23. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
24. Chomette, D., Frain, M., Cereghini, S., Charnay, P. & Ghislain, J. Krox20 hindbrain cis-regulatory landscape: interplay between multiple long-range initiation and autoregulatory elements. *Dev. Camb. Engl.* **133**, 1253–1262 (2006).
25. Ghislain, J. *et al.* Characterisation of cis-acting sequences reveals a biphasic, axon-dependant regulation of Krox20 during Schwann cell development. *Dev. Camb. Engl.* **129**, 155–166 (2002).
26. Faye, L. L., Machiela, M. J., Kraft, P., Bull, S. B. & Sun, L. Re-ranking sequencing variants in the post-GWAS era for accurate causal variant identification. *PLoS Genet.* **9**, e1003609 (2013).
27. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
28. Mendiola, M. *et al.* The orphan nuclear receptor DAX1 is up-regulated by the EWS/FLI1 oncoprotein and is highly expressed in Ewing tumors. *Int. J. Cancer J. Int. Cancer* **118**, 1381–1389 (2006).
29. Surdez, D. *et al.* Targeting the EWSR1-FLI1 oncogene-induced protein kinase PKC- $\beta$  abolishes ewing sarcoma growth. *Cancer Res.* **72**, 4494–4503 (2012).
30. Carrillo, J. *et al.* Cholecystokinin down-regulation by RNA interference impairs Ewing tumor growth. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **13**, 2429–2440 (2007).

**Main figures and figure legends:**
**Figure 1** Grunewald *et al.*

**Fig. 1: *EGR2* overexpression is mediated by EwS-specific eQTL and EWSR1-FLI1:**

a) *EGR2* and *ADO* expression levels in EwS (GSE34620) and normal tissues (GSE3526). Bars represent medians. Two-tailed unpaired student's t-test with Welch's correction.

b) Between-group analysis: Genes (grey dots) and tumour samples (spheres) are separated along three axes. EwS (n=279); RMS, rhabdomyosarcoma (n=121); OS, osteosarcoma (n=25); DSRCT, desmoplastic small-round-cell tumour (n=32); MB, medulloblastoma (n=52); NB, neuroblastoma (n=64); MRT, malignant rhabdoid tumour (n=35). c) eQTL analyses across tissue types identifies a EwS-specific correlation of *EGR2* and *ADO* expression with the risk-allele at rs1848797. The Broad GTEx database comprised 13 normal tissues with  $\geq 60$  samples/tissue type (total n=1421 samples). ns, not significant; AML, acute myeloid leukaemia; LCL, lymphoblastoid cell lines. d) Analysis of *EGR2* and *ADO* expression by qRT-PCR in human MSC lines L87 and V54-2 after ectopic EWSR1-FLI1 expression (pEWSR1-FLI1) as compared to empty vector (pControl). Mean and SEM,  $n \geq 9$ . The EWSR1-FLI1-targets *NR0B1* and *PRKCB* served as positive controls<sup>28,29</sup>. EWSR1-FLI1 expression was confirmed by Western blot (WB).

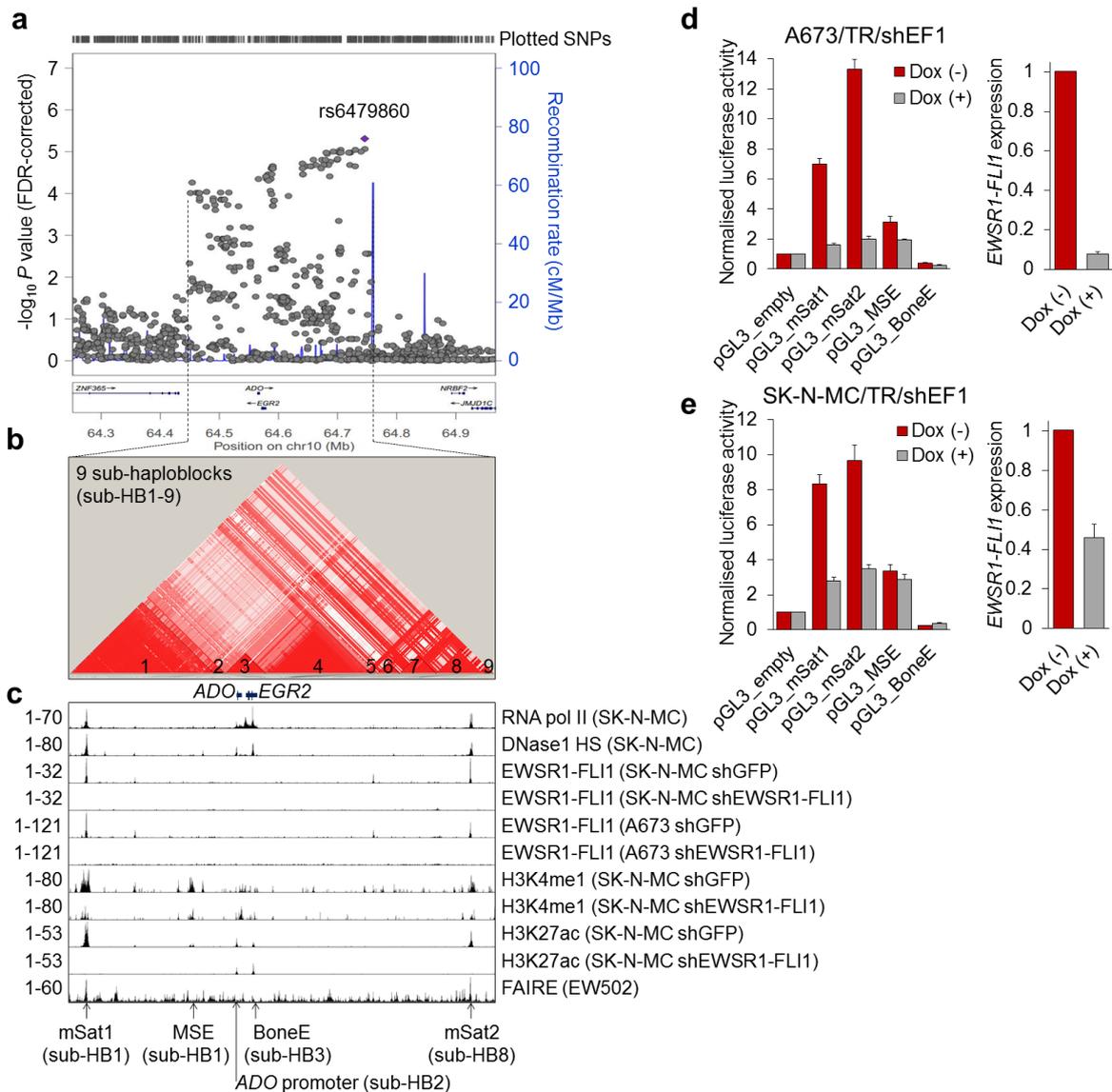
**Figure 2** Grunewald *et al.*



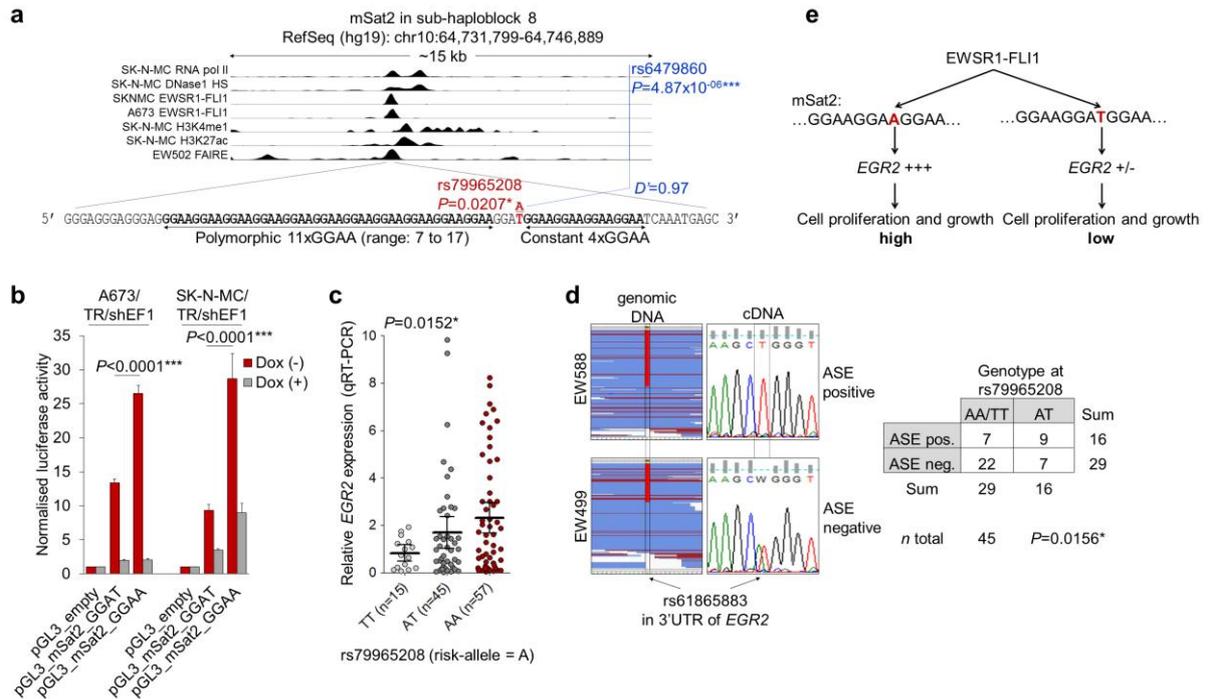
**Fig. 2: EGR2 is critical for growth and tumourigenicity of EwS:**

a) xCELLigence proliferation kinetics of A673 cells. Mean and SEM of results obtained with two different siRNAs against *EGR2* and three different siRNAs against *ADO*. Gene expression was measured by qRT-PCR at 48 h (mean and SEM,  $n \geq 4$ ). WB confirming *EGR2* knockdown (no suitable antibodies were available for *ADO*). b) Validation of xCELLigence results by cell counting (including supernatant) 96 h after transfection of the EwS cell lines A673, SK-N-MC, EW7, and POE (mean and SEM of results obtained with two different siRNAs against *EGR2* and three different siRNAs against *ADO*,  $n \geq 3$ ). c) Left: phase-contrast images of sphere-formation assays (scale bar = 1000  $\mu$ m). Right: Mean and SEM of  $n \geq 3$  experiments performed with SK-N-MC and POE cells containing a doxycycline-inducible shRNA against *EGR2* (shEGR2\_4 or shEGR2\_5). Similar results were obtained with A673 cells (not shown). Representative WB of *EGR2* in POE cells after 96 h of doxycycline-treatment. d) Growth curves (mean and SEM) of subcutaneously xenografted POE cells (shEGR2\_4). Doxycycline and sucrose (Dox +) or sucrose alone (Dox -) was added to the drinking water at day 12, when solid tumours were first palpable, and maintained until day 47. No doxycycline-effect was observed in xenografted POE cells containing an inducible non-targeting shRNA (not shown). Two-tailed unpaired student's t-test. \*\*\*  $P < 0.001$ . e) Venn diagrams of up- and downregulated genes 48 h after knockdown of *EWSR1-FLI1* or *EGR2* in A673 and SK-N-MC cells (min.  $\log_2$  fold-change  $\pm 0.5$ , BH-corrected  $P < 0.05$ ). Fisher's exact test.

**Figure 3** Grunewald *et al.*



**Fig. 3: Fine-mapping and epigenetic profiling reveals candidate *EGR2* regulatory elements:** a) Manhattan plot of 1466 SNPs identified by targeted deep-sequencing within the chr10 susceptibility locus. rs6479860 is the most significant SNP associated with EwS at this locus. The blue line indicates the recombination rate in the 1000 Genomes CEU population<sup>27</sup>. b) LD plot of the chr10 susceptibility locus hotspot (chr10:64,449,549-64,756,872) based on the analysis of 266 significantly EwS-associated SNPs in the entire cohort (n=600). c) Epigenetic profile of the chr10 susceptibility locus hotspot in the EwS cell lines SK-N-MC, A673, and EW502. Displayed are signals from published ChIP-seq or DNase-seq data for RNA polymerase II (pol II), DNase1 hypersensitivity (HS), as well as EWSR1-FLI1, H3K4me1, and H3K27ac in EwS cells either transfected with a control (shGFP) or specific shRNA (shEWSR1-FLI1), and FAIRE<sup>9,20,21</sup>. The read count is given on the left. mSat1 and mSat2: GGAA-microsatellites, MSE: Myelinating Schwann cell Enhancer, BoneE: Bone Enhancer. d,e) Normalized luciferase reporter signals in A673/TR/shEF1 and SK-N-MC/TR/shEF1 EwS cells containing a doxycycline-inducible shRNA against *EWSR1-FLI1*<sup>30</sup>. *EWSR1-FLI1* knockdown was confirmed by qRT-PCR. Mean and SEM, n≥5.

**Figure 4** Grunewald *et al.*


**Fig. 4: Germline variation at mSat2 modulates *EWSR1-FLI1*-dependant *EGR2* expression:** a) Epigenetic profile, genomic coordinates, and reference sequence of the mSat2 locus. Consistent with previous studies, signals for H3K4me1 and H3K27ac peak adjacent to the repetitive GGAA-microsatellite<sup>9,20</sup>. GGAA-repeats are underlined by arrows. The reported number of GGAA-motifs corresponds to the reference sequence. The *P* values for rs79965208 and rs6479680 reflect the significance of differences in genotype distribution between EwS and controls. b) Luciferase reporter signals of mSat2 as in (Fig 3d,e) with the T- or A-allele at rs79965208. Mean and SEM, n=4. Two-tailed unpaired student's t-test. c) *EGR2* expression measured by qRT-PCR in 117 EwS (103 primary tumours and 14 cell lines). Mean and 95%CI. The *P* value of a linear regression is reported. d) Representative Integrative Genomics Viewer pile-up of next-generation sequencing reads at the rs61865883 locus in genomic DNA and the corresponding electropherograms from Sanger sequencing of cDNA of the same primary EwS. The EwS sample EW588 exhibits ASE, whereas EW499 does not. The contingency table summarises the ASE results. One-tailed *Chi*<sup>2</sup> test. e) Model of the regulatory relationship of EWSR1-FLI1 and mSat2 controlling *EGR2* expression and consequently proliferation and growth of EwS cells.