Linking germline and somatic variation in Ewing sarcoma

Nicholas C Gomez & Ian J Davis

The identification of gene-regulatory polymorphisms that influence cancer susceptibility can identify key oncogenic pathways. A new study links a germline variant to Ewing sarcoma disease susceptibility and EWSR1-FLI1–mediated gene activation.

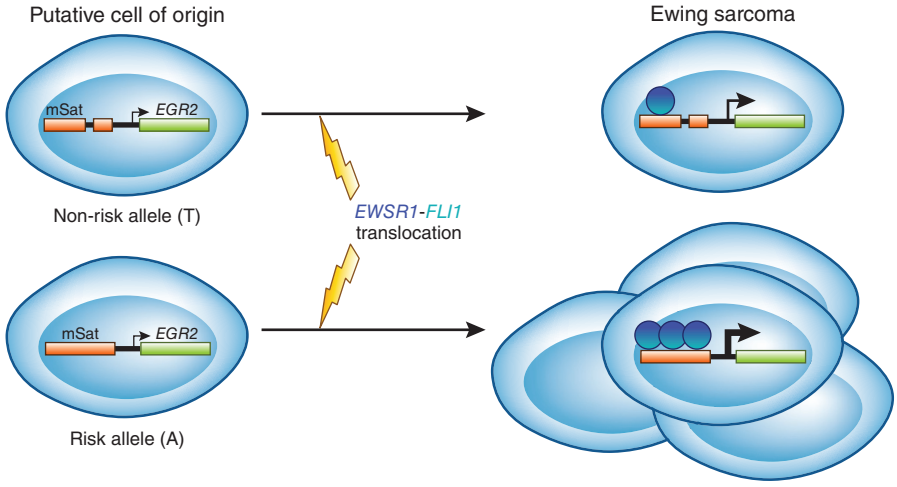
Cancers commonly demonstrate uneven demographic or familial distributions. The underpinnings of these differences often highlight specific genes or molecular pathways that have important roles. For many years, it has been appreciated that the malignant pediatric bone and soft tissue tumor Ewing sarcoma demonstrates a strong ancestry imbalance, with populations of European descent much more likely to be affected[1]. The molecular basis for this disparity remains unclear. Using a combination of high-throughput sequencing and integrative genomics, Olivier Delattre and colleagues hone in on an unusual segment of the genome that links germline variation with Ewing sarcoma risk[2].

Studies of Ewing sarcoma biology commonly focus on a critical chimeric oncoprotein that derives from the translocation-associated fusion of EWSR1 and FLI1. This oncoprotein acts as a potent modulator of gene expression by binding a microsatellite that consists of a repeated GGAA tetranucleotide motif[3–5]. Grünewald *et al.*[2] instead take a path less traveled and build upon the results of a Ewing sarcoma germline susceptibility study that identified SNPs on chromosome 10 as a risk locus[6]. Using integrative genomics, they have now identified a SNP in this region that increases EWSR1-FLI1 binding, thereby increasing expression of the transcription factor EGR2 (**Fig. 1**).

Collateral damage

The previous identification of a Ewing sarcoma susceptibility locus at 10q21.3 focused attention on *EGR2* and *ADO* as potential candidate genes[6]. Both genes are overexpressed in Ewing sarcoma relative to other normal tissues. Highlighting the unique importance of EGR2, Grünewald *et al.* demonstrate that only *EGR2* is dependent on EWSR1-FLI1 expression and that silencing of EGR2 results in decreased cell proliferation as well as decreased tumor growth, as assessed *in vitro* by spheroid formation and *in vivo* by mouse xenografts. In contrast, knockdown of ADO had no effect.

However, the real strength of the study derives from the targeted deep sequencing of the chromosome 10 susceptibility locus in 343 Ewing sarcoma cases and 251 genetically matched controls, which identified 291 common SNPs distributed among 10 subhaplo-blocks. The authors then prioritized SNPs for further study by examining publicly available genome-wide data sets, including data from formaldehyde-assisted isolation of regulatory elements (FAIRE) and DNase I hypersensitivity to explore chromatin accessibility, as well as chromatin immunoprecipitation to characterize EWSR1-FLI1 binding and histone post-translational modifications. These studies enabled them to focus on two microsatellites, each having chromatin organization and histone marks associated with active chromatin, bound by EWSR1-FLI1 and capable of conferring EWSR1-FLI1–mediated enhancer activity. One microsatellite consisted of a polymorphic GGAA repeat with a median series length of 11. This region was separated from another set of these repeats by only four nucleotides. Grünewald *et al.* found that a SNP encoding a T>A change in this linker region converted the GGAT motif into a GGAA sequence, which joined and extended the polymorphic repeat, ultimately resulting in a microsatellite with



Marina Corral Spence/Nature Publishing Group

Figure 1 Inherited variation on chromosome 10 near *EGR2* influences Ewing sarcoma susceptibility by altering a binding site for EWSR1-FLI1. By extending a microsatellite, a common germline variant facilitates binding of the EWSR1-FLI1 chimeric oncoprotein to the *EGR2* locus, leading to higher *EGR2* expression and increased susceptibility to Ewing sarcoma. mSat, microsatellite.

Nicholas C. Gomez and Ian J. Davis are in the Department of Genetics and the Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
e-mail: ian_davis@med.unc.edu

~16 repeats. The authors proceeded to show that the A-containing risk allele exhibited enhanced luciferase activity *in vitro* and was associated with increased *EGR2* expression in a cohort of 103 primary tumors and 14 Ewing sarcoma cell lines. Interestingly, the risk allele is found at significantly higher frequencies in non-African populations.

Although varying the number of binding motifs from 11 to 16 may seem unimpressive, studies have shown the relevance of microsatellite length to EWSR1-FLI1 binding and responsiveness[3]. From a genomics perspective, EWSR1-FLI1 binding *in vivo* is most enriched for contiguous lengths of 12–15 motifs, with enrichment decreasing rapidly at longer microsatellites[4]. Functionally, repeats of 20 motifs demonstrate far greater EWSR1-FLI1–mediated activation than repeats that are much smaller or larger. Interestingly, repeats with 50 motifs again become responsive to EWSR1-FLI1, although with somewhat attenuated activity[7]. In addition, contiguous repeats demonstrate greater activity than those that are disrupted. Other microsatellite regions demonstrate length polymorphisms, including the regions near *NR0B1*, *CAV1* and *IGF1*, genes important for Ewing sarcoma development and treatment[4,8]. Monument *et al.* explored the length of the *NR0B1* microsatellite and found longer repeats in tumors than in matched controls[7]. Taken together, these data all point to the importance of microsatellite length as a factor influencing Ewing sarcomagenesis.

Microsatellites and beyond

This study highlights the ability to narrow the search for functional variants by informing genome-wide association studies (GWAS) with genome-wide chromatin data, as the functional variant was not the most disease-associated SNP in the haploblock. Although SNPs affecting transcription factor binding to regulatory elements and consequent gene expression have been identified, localizing the functional SNP to a microsatellite again highlights the importance of these regions to Ewing sarcoma development. This study also expands our appreciation of the ways that germline enhancer variation can contribute to cancer risk.

The variation in microsatellite length between populations identified in this study suggests a potential mechanism for the ancestry differences seen in Ewing sarcoma that have perplexed clinicians for years. In the absence of an apparent familial predisposition, however, it is likely that germline variants have a subtle role in tumor formation. It is possible that the distribution of this SNP within the population, in combination with a low rate of chromosomal rearrangement, limits the impact of the SNP. Although not explored in the current study, microsatellite differences may also have prognostic value. Microsatellite variations may be genetically linked, such that inherited combination of microsatellite variants affects both susceptibility and response to treatment.

This study again demonstrates the importance of EWSR1-FLI1 targeting to microsatellites for gene activation. Although repetitive elements are found in other species, the classes and relationships with proximal genes differ, offering a possible explanation for the long-standing challenge in generating an animal model for Ewing sarcoma.

We lack a full understanding of the nature of the elements in the cell of origin permitting targeting by EWSR1-FLI1. In fact, the cell of origin remains a contentious topic, with evidence implicating both neural-derived precursors and mesodermally derived mesenchymal stem cells. Future studies shedding light on the features that allow classes of repetitive elements to be co-opted by EWSR1-FLI1 may further understanding of this cancer, with hope for targeted therapies and better prognostics for children and young adults with this devastating disease.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Miller, R.W. *Recent Results Cancer Res.* **1976**, 50–62 (1976).
2. Grünewald, T.G.P. *et al. Nat. Genet.* **47**, 1073–1078 (2015).
3. Gangwal, K. *et al. Proc. Natl. Acad. Sci. USA* **105**, 10149–10154 (2008).
4. Patel, M. *et al. Genome Res.* **22**, 259–270 (2012).
5. Riggi, N. *et al. Cancer Cell* **26**, 668–681 (2014).
6. Postel-Vinay, S. *et al. Nat. Genet.* **44**, 323–327 (2012).
7. Monument, M.J. *et al. PLoS ONE* **9**, e104378 (2014).
8. Beck, R. *et al. Cancer Genet.* **205**, 304–312 (2012).

Genetic differential calculus

Richard Mott

High-throughput analysis of the phenotypes of mouse genetic knockouts presents several challenges, such as systematic measurement biases that can vary with time. A report from the EUMODIC consortium presents data from 320 genetic knockouts generated using standardized phenotyping pipelines and new statistical analyses aimed at increasing reproducibility across centers.

What is the function of a gene? This is an ambiguous question, as the answer depends on which aspects of the organism's phenotype are investigated, the genetic background and the environment. At present, we cannot reliably predict function computationally, so it has to be determined experimentally, by altering a gene and observing the consequences. Systematically generated catalogs of gene function based on data from knocking out each gene separately are conceptually attractive and are underway in many model organisms. A report by Steve Brown and colleagues[1] describes the state of the art in the mouse.

Richard Mott is at the Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
e-mail: rmott@well.ox.ac.uk

The challenges of high-throughput phenotyping

The study represents a significant experimental effort. To ensure consistent phenotyping, hundreds of gene knockout lines were generated, thousands of animals were bred and maintained in clean and uniform conditions, prospective phenotypes were measured consistently over time and between institutes, and the data were analyzed to take account of confounding sources of variation. Finally, the data and annotations had to be made freely available as transparently as possible. The work was spread across four centers in the UK, France and Germany that are part of the EUMODIC consortium, and the data were analyzed and released by Medical Research Council (MRC)-Harwell, UK.