Functional annotation of susceptibility loci identified by COGS

It is likely that many of the cancer risk loci found by COGS influence gene regulation, as ~66% of confirmed susceptibility alleles lie within or near genes, with only 5% located within coding regions. Subsequent functional annotation and experimental studies therefore have sought spatial association with transcription factor binding sites, histone modifications or other putative regulatory elements. However, *in vitro* functional studies in a relevant tissue and cell type, as well as *in vivo* studies in model organisms, are also needed to provide insight into the mechanisms underlying the development of breast, ovarian and prostate cancers.

Orli Bahcall

Framework for functional analysis

Genome-wide association studies (GWAS) have identified hundreds of loci that are associated with susceptibility to a wide range of common diseases and traits (*N. Engl. J. Med.* **363**, 166–176, 2010). For almost all identified common susceptibility loci for cancer, the functional basis underlying disease risk remains unknown. This may in part reflect the diversity of the biological mechanisms that underlie cancer predisposition but also the challenges in moving from association studies to function. One challenge is that over 90% of the common genetic trait-associated variations identified by GWAS lie in noncoding DNA regions (*Science* **337**, 1190–1195, 2012), and annotation of the functional role of these regions remains limited.

In addition to the identification of genetic associations, the papers reported in this COGS collection also establish a framework for the preliminary functional analysis of common, low-penetrance cancer susceptibility loci (**Fig. 1**) (*Nat. Genet.* **43**, 513–518, 2011). Because over 95% of the identified susceptibility variants are found within noncoding regions, these studies have mainly focused on the annotation of risk-associated SNPs with respect to noncoding regulatory elements linked to the identification of plausible candidate susceptibility genes. This was done using data generated in large collaborative initiatives including the Encyclopedia of DNA Elements (ENCODE), the 1000 Genomes Project, The Cancer Genome Atlas (TCGA) and the International Cancer Genome Project. The current studies include initial efforts at functional annotation, and further studies will be needed that are based on more comprehensive fine-mapping efforts and experimental studies, including functional assays and *in vitro* and *in vivo* modeling of candidate susceptibility genes in disease development.

Location, location, location

Even though less than 5% of identified associated variants in the COGS studies are exonic, approximately two-thirds of all confirmed susceptibility alleles are intragenic or in close proximity to a candidate gene, suggesting possible gene targets at susceptibility loci. In the absence of fine mapping and detailed functional analyses, caution needs to be applied in interpreting the significance of candidate genes spanning or close to associated variants in susceptibility loci. However, it is worth noting the presence of plausible cancer susceptibility genes, particularly when present across cancer types.

Examples from the papers in the COGS collection include the *TERT* gene at 5p15.33, with SNPs reported to be associated with breast, prostate and ovarian cancers, low-malignant-potential (LMP) ovarian cancer and telomere length. SNPs in or near the *MDM4* gene at the 1q32 locus were associated with breast and prostate cancers and were identified as modifiers of the risk of breast cancer in *BRCA1* mutation carriers. SNPs in or near

the *HNF1B* gene at the 17q12 locus were associated with prostate cancer and two different subtypes of ovarian cancer. SNPs in or near *RAD51B* at the 14q24 locus were associated with prostate cancer.

Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array

Eeles, R.A., Al Olama, A.A., Benlloch, S., Saunders, E.J., Leongamornlert, D.A. *et al.*

doi:10.1038/ng.2560

All of the newly associated loci lie in linkage disequilibrium (LD) blocks that include plausible causative genes (**Fig. 3a–d** and **Supplementary Fig. 3**).

Fifteen of the 23 SNPs are either intronic (12 SNPs) or in the promoter region of a gene (3 SNPs).

rs4245739 at 1q32 is situated in the 3' UTR of the *MDM4* gene, 32 bp downstream of the stop codon.

rs11568818 at 11q22 lies within a small LD region containing a single gene, *MMP7*, encoding a matrix metalloproteinase.

rs7141529 at 14q24 lies within the last intron of the longest isoform of *RAD51B* (also known as *RAD51L1*). Members of the RAD51 family are involved in the repair of double-stranded DNA breaks by homologous recombination, and their loss is potentially oncogenic.

rs11650494 is located at 17q21, a gene-dense locus that contains several genes that have been proposed as potential prostate cancer susceptibility or somatically altered genes, including *HOXB13*, *PRAC*, *SPOP* and *ZNF652*.

Large-scale genotyping identifies 41 new loci associated with breast cancer risk

Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J. *et al.*

doi:10.1038/ng.2563

Two associated loci lie within or close to known breast cancer susceptibility genes. rs11571833 is a polymorphic variant in *BRCA2* that introduces a premature stop codon (p.Lys3326*), previously reported to have no association with breast cancer risk²⁷. The results from the current study, however, indicate that this variant is associated with a modestly higher risk of breast cancer.

In addition to rs11571833, one further SNP is a coding variant: rs11552449 encodes a missense substitution p.His61Tyr in *DCLRE1B* (also known as *SNM1B*), an evolutionarily conserved gene involved in DNA stability and the repair of interstrand cross-links²⁹. The remaining loci are either intronic (20) or intergenic (19). Two loci lie within genes previously proposed as candidate breast cancer susceptibility genes. SNP rs12493607 lies in intron 2 of *TGFB2*.

Multiple independent variants at the *TERT* locus are associated with telomere length and risk of breast and ovarian cancer

Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K. *et al.*

doi:10.1038/ng.2566

Correlated SNPs in the *TERT* promoter (peak 1) were associated with telomere length, ER-positive breast cancer, ER-negative breast cancer and breast cancer in *BRCA1* mutation carriers. SNPs in peak 2, spanning *TERT* introns 2–4, were independently associated with telomere length, overall breast cancer risk and serous LMP ovarian cancer. SNPs in peak 3, also spanning *TERT* introns 2–4, showed strong associations with ER-negative breast cancer, breast cancer risk for *BRCA1* mutation carriers and serous invasive ovarian cancer but not with telomere length (**Tables 1 and 2**).

Use public databases

Another useful approach is the use of publicly available data from genome profiling studies of tumor tissues. The examples noted below have made use of data sets to study differential expression in cancer compared to normal tissues, DNA copy number variation and differential patterns of DNA methylation. These studies have provided evidence of CpG island methylation of the *TERT* and *CLPTM1L* genes at 5p15.33 and of *HNF1B* at 17q12 in ovarian cancer. Several of these comparisons have provided support for the involvement of the candidate genes *C10orf114* and *MLLT10* at 10q12 and *ARHGAP27* and *PLEKHM1* at 17q21.31 in ovarian cancer susceptibility.

Multiple independent variants at the *TERT* locus are associated with telomere length and risks of breast and ovarian cancer

Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K. *et al.*

doi:10.1038/ng.2566

We used The Cancer Genome Atlas (TCGA)⁴⁹ data to examine gene expression of the 11 protein-coding genes and 1 microRNA (*MIR4457*) located within 1 Mb of peak 3 SNP rs10069690. Most genes showed higher expression in ovarian tumors compared with normal tissues (**Supplementary Fig. 4** and **Supplementary Table 7**). We observed no association between rs10069690 and the expression levels of any of the genes in any of the cells tested (**Supplementary Fig. 5** and **Supplementary Tables 7 and 8**). There was some evidence of association between rs10069690 and tumor methylation with probes cg23827991 (*TERT* CpG island, $P = 1.3 \times 10^{-6}$) and cg06550200 (*CLPTM1L*, $P = 6.9 \times 10^{-4}$) out of the 935 probes tested. Both regions showed lower methylation with the minor, cancer risk-associated allele (**Supplementary Table 9**), but this did not correlate with changes in expression.

GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer

Pharoah, P.D.P., Tsai, Y.-Y., Ramus, S.J., Phelan, C.M., Goode, E.L. *et al.*

doi:10.1038/ng.2564

Correlations between gene expression and DNA copy number variation at this locus in primary EOC [epithelial ovarian cancer] tissues suggest that overexpression of *C10orf114* and *MLLT10* is driven by copy number variation.

Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31

Permuth-Wey, J. *et al.*

doi:10.1038/ncomms2613

rs12942666 and many of its correlated SNPs lie within introns of Rho GTPase activating protein 27 (*ARHGAP27*) or its neighboring gene, pleckstrin homology domain containing, family M (with RUN domain)

member 1 (*PLEKHM1*) (**Supplementary Table S2**). There are another 15 known protein-coding genes within the region: *KIF18B*, *C1QL1*, *DCAKD*, *NMT1*, *PLCD3*, *ABCB4*, *HEXIM1*, *HEXIM2*, *FMNL1*, *C17orf46*, *MAP3K14*, *C17orf69*, *CRHR1*, *IMP5*, and *MAPT* (**Fig. 2a**).

To evaluate the likelihood that one or more genes within this region represent target susceptibility gene(s), we first analyzed expression and methylation of these genes in EOC tissues and cell lines (**Fig. 2b–g** and **Supplementary Tables S3** and **S4**). Most genes showed significantly higher expression ($P < 10^{-4}$) in EOC cell lines versus normal ovarian cancer precursor tissues (OCPTs); *ARHGAP27* showed the most pronounced difference in gene expression between cancer and normal cells ($P = 10^{-16}$) (**Fig. 2b** and **Supplementary Table S3**). For nine genes, we also found overexpression in primary high-grade serous (HGS) EOC tumors versus normal ovarian tissue in at least one of two publicly available datasets, The Cancer Genome Atlas (TCGA) of 568 tumors and/or the Gene Expression Omnibus (GEO) series GSE18520 dataset consisting of 53 tumors¹⁶ (**Fig. 2c** and **Supplementary Table S3**). Analysis of DNA copy number variation in TCGA revealed frequent loss of heterozygosity in this region (**Supplementary Fig. 5a**; **Supplementary Methods**), and correlations with gene expression suggested alternate mechanisms (such as epigenetics) may be driving the observed overexpression. We observed significant hypomethylation ($P < 0.01$) in ovarian tumors for *DCAKD*, *PLCD3*, *ACBD4*, *FMNL1*, and *PLEKHM1* (**Fig. 2d** and **Supplementary Table S4**), which is consistent with the overexpression observed for *DCAKD*, *PLCD3*, and *FMNL1*.

eQTL analysis

Quantitative expression trait (eQTL) analysis can be used to search for genes differentially expressed at susceptibility loci by testing for correlation of the disease-associated SNP with variants transcribed into RNA. eQTL analyses are ideally performed in normal tissues, where the tissue samples genotyped are diploid, but can also be performed in tumor tissues after adjusting for the effects of somatic copy number and methylation. A variation on this approach is to examine correlations between methylation at CpG sites and genotype (mQTL). Critically, gene expression and methylation can be highly tissue specific, and, therefore, normal eQTL and mQTL analyses need to be performed in the tissues of origin for the phenotype under study. The examples below highlight how eQTL and mQTL analyses were used as functional assays in the COGS studies. Multiple risk SNP–gene interactions, suggesting possible target genes and functional associated alleles, were identified, including associations between rs616402 and *PEX14* expression in breast cancer; rs11782652 and *CHMP4C* in ovarian cancer; rs9348512, a *BRCA2*-modifying allele, and *GCNT2* expression in breast tumors; rs2077606 and *PLEKHM1* expression in ovarian cancer; and rs11658063 and *HNF1B* methylation in ovarian cancer.

Large-scale genotyping identifies 41 new loci associated with breast cancer risk

Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J. *et al.*

doi:10.1038/ng.2563

To further investigate the likely genes underlying the susceptibility variants, we examined associations between the lead SNPs and the RNA expression of neighboring genes in 473 primary breast tumors and 61 normal breast tissue samples in The Cancer Genome Atlas (TCGA) database. We found strong evidence for an association between rs616402 (a surrogate for rs616488; $r^2 = 0.66$) and expression of *PEX14* in both tumor ($P = 4.7 \times 10^{-12}$) and normal tissue ($P = 0.00018$; **Supplementary Table 8**), between rs3760983 (a surrogate for rs3760982; $r^2 = 1$) and expression of both *ZNF404* ($P = 1.2 \times 10^{-6}$ in tumors) and *ZNF283* ($P = 0.0089$) and between rs3903072 and expression of *CTSW* ($P = 4.9 \times 10^{-5}$). SNP rs3760982 was also found to be associated with the expression of *ZNF45* ($P = 0.0077$), *ZNF283* ($P = 0.05$) and *ZNF222* ($P = 0.01$) in lymphoblastoid cell line from HapMap samples using the Genevar database³⁵ (**Supplementary Table 8c**). After adjustment for the SNP in the region most strongly associated with expression, SNP rs616488 and *PEX14* ($P =$

Figure 1 | Summary of a pipeline for functional annotation of common variants at cancer susceptibility loci. After the identification of common variants associated with disease, fine mapping of the susceptibility loci is a required step in resolving likely causal allele(s). Integrating fine-mapping data with genome-wide analysis of noncoding regulatory elements may be used to identify the putative functional targets of trait-associated alleles. Performing expression quantitative trait locus (eQTL) analysis in normal and/or cancer tissues can be used to link susceptibility SNPs overlapping regulatory elements with candidate susceptibility genes. Finally, a combination of *in vitro* and *in vivo* functional assays can be used to characterize the mechanisms and roles in disease etiology of both candidate regulatory elements and candidate genes.

0.0071) as well as rs1217396 (a proxy for rs11552449) and *PTPN22* ($P = 0.0055$) and *DCLRE1B* ($P = 0.0067$) reached nominal significance at $P < 0.01$ (**Supplementary Table 8a**). Although none of these passed Bonferroni correction for multiple testing, the three associations found exceeded the number expected by chance with 46 associations tested. This supports some transcriptional effect from the risk-associated SNPs. *PEX14* is involved in peroxisome organization and protein and transmembrane transport; mutations in *PEX14* have been associated with Zellweger syndrome³⁶. The functions of ZNF45, ZNF222 and ZNF283 are unknown but may involve transcriptional regulation.

GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer

Pharoah, P.D.P., Tsai, Y.-Y., Ramus, S.J., Phelan, C.M., Goode, E.L. *et al.*

doi:10.1038/ng.2564

We found no evidence of a correlation between rs11782652 genotype and gene expression in normal ovarian or fallopian tube epithelial cells for any of the nine genes in the region (*FABP5*, *PMP2*, *FABP4*, *FABP12*, *IMPA1*, *SLC10A5*, *ZFAND1*, *CHMP4C* and *SNX16*), but there was a highly statistically significant association between rs11782652 and *CHMP4C* expression in primary EOC tissues ($P = 3.9 \times 10^{-14}$) and transformed lymphocytes ($P = 0.012$).

Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31

Permuth-Wey, J. *et al.*

doi:10.1038/ncomms2613

We evaluated associations between genotypes for the top risk SNP rs12942666 (or a tagSNP) and expression of all genes in the region (expression quantitative trait locus (eQTL) analysis) in normal OCPTs, lymphoblastoid cell lines and primary ovarian tumours from TCGA. The only significant eQTL association observed ($P < 0.05$) in normal OCPTs was for ARHGAP27 ($P = 0.04$) (**Fig. 2e; Supplementary Table S3**). Because rs12942666 was not genotyped in tissues analysed in TCGA, we used data for its correlated SNP rs2077606 ($r^2 = 0.99$) to evaluate eQTLs in tumour tissues. Rs2077606 genotypes were strongly associated with PLEKHM1 expression in primary HGS-EOCs ($P = 1 \times 10^{-4}$) (**Fig. 2f; Supplementary Table S3**).

Epigenetic analysis leads to identification of *HNF1B* as a subtype-specific susceptibility gene for ovarian cancer

Shen, H. *et al.*

doi:10.1038/ncomms2629

We further investigated the relationship between risk allele genotypes and *HNF1B* DNA methylation in 231 serous ovarian cancers. The top serous risk SNP, rs7405776, showed only a borderline association with increased promoter methylation ($P = 0.07$; **Fig. 3**). Intriguingly, the association between SNPs in *HNF1B* and *HNF1B*-promoter DNA methylation strengthened as their location approached the promoter region, and the strongest signal came from a few SNPs, exemplified by rs11658063, overlapping with a polycomb repressive complex 2 (PRC2) mark in embryonic stem cells ($P = 0.003$; **Fig. 3, Supplementary Fig. S8**).

Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with *TERT* expression

Kote-Jarai, Z. *et al.*

doi:10.1093/hmg/ddt086

They examined gene expression of *TERT* and *CLPTM1L* in 195 normal (histologically benign) prostate tissue samples isolated from men with elevated prostate-specific antigen (PSA) levels. They found protective alleles of four SNPs in one region associated with higher expression of *TERT*.

Research Highlight in *Nature Genetics*, doi:10.1038/ng.2597

Regulatory elements

Using a combination of data from ENCODE and profiling methods for DNA-protein complexes such as formaldehyde-assisted investigation of regulatory elements with sequencing (FAIRE-seq) or chromatin immunoprecipitation with sequencing (ChIP-seq), several studies in the COGS collection were able to identify evidence of overlap between susceptibility SNPs and putative regulatory elements and transcription factor binding sites. However, this is only part of the challenge; it is also critical to demonstrate that a putative regulatory element is active in the tissue of origin, as shown at 11q13, where a strong transcriptional enhancer led to an increase in *CCND1* promoter activity.

GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer

Pharoah, P.D.P., Tsai, Y.-Y., Ramus, S.J., Phelan, C.M., Goode, E.L. *et al.*

doi:10.1038/ng.2564

Encyclopedia of DNA Elements (ENCODE) data from tissues not associated with ovarian cancer, formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) data and mapping of enhancer elements generated in normal serous ovarian cancer precursor cells suggested that there are two regulatory regions that may be influenced by risk-associated SNPs: one at the *CHMP4C* promoter and the other in intron 1 of *CHMP4C* (**Fig. 2**).

Multiple independent variants at the *TERT* locus are associated with telomere length and risks of breast and ovarian cancer

Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K. *et al.*

doi:10.1038/ng.2566

Analysis of Encyclopedia of DNA Elements (ENCODE) data³⁹ showed no evidence of regulatory elements or open chromatin coinciding with any risk-associated SNPs in normal breast epithelial cells or the other represented tissues (**Supplementary Fig. 4**). Data for ovarian tissues are not included in ENCODE. We therefore performed site-specific formaldehyde-assisted isolation of regulatory elements (FAIRE)⁴⁰ in ovarian cancer precursor tissues to identify regulatory elements in a 1 Mb region centered on peak 3. In fallopian tube secretory and ovarian surface epithelial cells, we detected FAIRE peaks coinciding with the *CLPTM1L* promoter but not the *TERT* promoter (**Supplementary Fig. 4**). *In silico* analyses additionally indicated that *TERT* introns 4 and 5 (within and beyond peak 3) contained regions showing regulatory potential and vertebrate sequence conservation⁴¹. We performed site-specific FAIRE analyses on a ~1 kb region centered on the peak 3 SNP rs10069690 in normal tissue samples from breast reduction mastoplasty ($n = 4$), ovarian cancer precursor tissues ($n = 4$) and ovarian cancer cell lines ($n = 4$). Breast cells from each woman were sorted into four enriched fractions on the basis of differential expression of cell surface markers⁴² (myoepithelial/stem, luminal progenitor, mature luminal and stromal cells), and assays were performed on each fraction (**Fig. 3**). Chromatin was in a closed configuration in all ovarian, breast luminal progenitor and mature luminal cell fractions. However, in two of four stromal cell fractions, we detected ~600 bp of open chromatin of varying amplitude, covering the position of SNP

rs10069690 but not of rs2242652, and, in three of four myoepithelial/stem cell fractions, we detected ~800 bp of open chromatin, covering the positions of both SNPs rs10069690 and rs2242652.

Genome-wide association studies identify four ER negative-specific breast cancer risk loci

Garcia-Closas, M., Couch, F.J., Lindsrtom, S., Michailidou, K., Schmidt, M.K. *et al.*

doi:10.1038/ng.2561

The signal found on chromosome 16q12.2 is located in the fat mass- and obesity-associated gene *FTO* (**Supplementary Fig. 4d**). This signal is tagged by rs11075995, located in a ~40-kb LD block in intron 1 of *FTO*, within an enhancer region that appears to be active in both normal and triple-negative breast cancer cells (**Supplementary Fig. 5d**).

Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array

Eeles, R.A., Al Olama, A.A., Benlloch, S., Saunders, E.J., Leongamornlert, D.A. *et al.*

doi:10.1038/ng.2560

rs4245739 has been shown to create an illegitimate binding site for miR-191 that results in the downregulation of *MDM4* expression¹²; this is in agreement with our analysis using mirsnpscore¹³, which predicted that the risk allele creates a binding site for miR-191, miR-887 and miR-3669. However, rs4245739 is also highly correlated with a number of other *MDM4* variants that overlap functional elements identified by the Encyclopedia of DNA Elements (ENCODE) Project^{13,14}. Other analyses using the iCOGS array have found that rs4245739 and correlated SNPs are associated with estrogen receptor (ER)-negative breast cancer¹⁵ and breast cancer in *BRCA1* mutation carriers¹⁶. In addition, the risk allele (C) of rs4245739 has been associated with increased aggressiveness in individuals with ovarian cancer¹⁷.

rs11568818 at 11q22 lies within a small LD region containing a single gene, *MMP7*, encoding a matrix metalloproteinase. Matrix metalloproteinases are implicated in metastasis, and elevated *MMP7* expression itself has been reported as a potential biomarker for metastatic prostate cancer and poor disease prognosis¹⁸. This SNP is situated 181 bp upstream of the transcriptional start site in the promoter region, within an area of high sequence conservation that overlaps strong DNase hypersensitivity and transcription factor binding sites^{13,14}. rs11568818 itself has been established as a functional promoter variant, with the risk allele (A) having been shown to create a binding site for the FOXA2 transcription factor and result in higher *MMP7* expression¹⁹.

Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31

Permuth-Wey, J. *et al.*

doi:10.1038/ncomms2613

To explore the possible functional significance of rs12942666 and strongly correlated variants ($r^2 > 0.80$), we then generated a map of regulatory elements around rs12942666 using ENCYClopedia of DNA Elements (ENCODE) data and formaldehyde-assisted isolation of regulatory elements sequencing analysis of OCPTs (**Supplementary Methods**). We observed no evidence of putative regulatory elements coinciding with rs12942666 or correlated SNPs (**Fig. 3a**). A map of regulatory elements in the entire 1-MB region can be seen in **Supplementary Fig. 5c–f**. We subsequently used *in silico* tools (ANNOVAR²³, SNPinfo²⁴ and SNPnexus²⁵) to evaluate the putative function of possible causal SNPs (**Supplementary Methods**). Of 50 SNPs with possible functional roles, more than 30 reside in putative transcription factor binding sites (TFBS) within or near

PLEKHM1 or *ARHGAP27*; 12 SNPs may affect methylation or miRNA binding, and two are non-synonymous coding variants predicted to be of no functional significance (**Supplementary Table S2**).

Further examples of functional annotation of regulatory elements can be found in Research Highlights of the papers by French *et al.* and Kote-Jarai *et al.*

Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers
French, J.D. *et al.*

doi:10.1016/j.ajhg.2013.01.002

They selected five promising candidate SNPs for functional studies but did not detect any significant association of these SNPs with the expression of local genes in normal breast tissue or tumor samples. Using chromatin immunoprecipitation with sequencing (ChIP-seq) data from MCF7 cells, they found that these SNPs fell within two putative regulatory elements. Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and chromosome conformation capture (3C) analyses showed long-range interactions between these regulatory elements and the *CCND1* promoter and/or terminator. Further functional studies identified a candidate causal variant in the putative *CCND1* enhancer, which affected the binding of the ELK4 transcription factor. A second candidate causal variant, located within a silencer element that physically interacts with the *CCND1* enhancer, affected binding of the GATA3 transcription factor.

Research Highlight in *Nature Genetics*, doi:10.1038/ng.2596