



Published in final edited form as:

Nature. 2011 February 10; 470(7333): 264–268. doi:10.1038/nature09753.

## 9p21 DNA variants associated with Coronary Artery Disease impair IFN $\gamma$ signaling response

Olivier Harismendy<sup>1,2,%</sup>, Dimple Notani<sup>3,%</sup>, Xiaoyuan Song<sup>3</sup>, Nazli G. Rahim<sup>4</sup>, Bogdan Tanasa<sup>3</sup>, Nathaniel Heintzman<sup>5</sup>, Bing Ren<sup>5</sup>, Xiang-Dong Fu<sup>6</sup>, Eric J. Topol<sup>4,\*</sup>, Michael G. Rosenfeld<sup>3,\*</sup>, and Kelly A. Frazer<sup>1,2,7,\*</sup>

<sup>1</sup> Department of Pediatrics and Rady's Children's Hospital, University of California at San Diego, School of Medicine, La Jolla, California 92093

<sup>2</sup> Moores UCSD Cancer Center, University of California San Diego, La Jolla California 92093

<sup>3</sup> Department of Medicine, Howard Hughes Medical Institute, University of California at San Diego, School of Medicine, La Jolla, California 92093

<sup>4</sup> Scripps Genomic Medicine, Scripps Translational Science Institute, The Scripps Research Institute, 3344 N. Torrey Pines Court, La Jolla CA 92037

<sup>5</sup> Ludwig Institute for Cancer Research - University of California at San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093-0653

<sup>6</sup> Department of Cellular and Molecular Medicine, University of California at San Diego, School of Medicine, La Jolla, California 92093

<sup>7</sup> Institute for Genomic Medicine, University of California San Diego, La Jolla California, 92093

### Abstract

Genome wide association studies (GWAS) have identified SNPs in the 9p21 gene desert associated with coronary artery disease (CAD)<sup>1–4</sup> and Type 2 diabetes (T2D)<sup>5–7</sup>. Despite evidence for a role of the associated interval in neighboring gene regulation<sup>8–10</sup>, the biological underpinnings of these genetic associations to CAD or T2D have not yet been explained. Here we identify 33 enhancers in 9p21; the interval is the second densest gene-desert for predicted enhancers and 6 times denser than the whole genome ( $p < 6.55 \times 10^{-33}$ ). The CAD risk alleles of SNPs rs10811656/rs10757278 are located in one of these enhancers and disrupt a binding site for STAT1. Lymphoblastoid cell lines (LCL) homozygous for the CAD risk haplotype exhibit no binding of STAT1, and in LCL homozygous for the CAD non-risk haplotype binding of STAT1 inhibits *CDKN2BAS* expression, which is reversed by siRNA knock-down of *STAT1*. Using a new, open-ended approach to detect long-distance interactions (3D-DSL), we find that in human vascular endothelium cells (HUVEC) the enhancer interval containing the CAD locus physically interacts with the *CDKN2A/B* locus, the *MTAP* gene and an interval downstream of *INFA21*. In HUVEC, IFN $\gamma$  activation strongly affects the structure of the chromatin and the transcriptional regulation in the 9p21 locus, including STAT1 binding, long-range enhancer interactions and

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Corresponding Authors: etopol@scripps.edu; mrosenfeld@ucsd.edu; kafrazer@ucsd.edu.

%These authors made equal contributions

#### Authors Contributions

OH, DN, BR, EJT, KAF and MGR designed the experiments. DN, XS, NGR, NH and XDF carried out the experiments. OH, DN, XS, BT, KAF and MGR contributed to analyzing the data. OH, DN, KAF and MGR wrote the manuscript.

**Author Information** The authors declare no competing interests

altered expression of neighboring genes. Our findings establish a link between CAD genetic susceptibility and the response to inflammatory signaling in a vascular cell type and thus demonstrate the utility of GWAS findings to direct studies to novel genomic loci and biological processes important for disease etiology.

---

GWAS have identified 8 SNPs in the 9p21 interval strongly associated with CAD<sup>1-4</sup> and other vascular diseases<sup>11,12</sup>, all of which are highly correlated ( $r^2 > 0.8$ ) and located in a 53-kb linkage disequilibrium (LD) block. The haplotype diversity in this interval is limited in Caucasians with ~ 25% of individuals homozygous for the risk haplotype which confers ~ 2 fold greater risk of myocardial infarction than noncarriers<sup>2</sup>. Independent studies have identified 4 more SNPs associated with T2D in an adjacent but yet distinct 11-kb LD block<sup>5-7</sup>. The associated CAD and T2D SNPs (Supplemental Table 1) lie in a gene desert (Figure 1) flanked by *CDKN2B* (130kb upstream) and *DMRTA1* (370kb downstream) suggesting that the functional variants underlying the association are likely in regulatory elements. The CAD interval overlaps with the 3' end of a non-coding gene with unknown function, *CDKN2BAS*, which is co-expressed with the *CDKN2A/B* locus<sup>13</sup>.

In this study we used a multi-pronged approach involving cellular assays and population sequencing to identify and functionally characterize the variants underlying the 9p21 association with CAD (Supplemental Figure 1). We initially sought to identify regulatory elements in the 9p21 gene desert by examining transcription factor binding and chromatin modification profiles in human cells including HeLa, K562, and human ES cells<sup>14</sup>. Histone H3 trimethylated at lysine4 (H3K4me3) is associated with promoters of active genes, looking for this mark we determined that the *CDKN2B* and *DMRTA1* promoters were the only ones in the interval (Figure 1). CTCF-binding sites mark insulators<sup>15</sup>; from ChIP-chip data of this factor in HeLa cells, we identified seven potential insulators in the 9p21 interval. One insulator is located close to the *CDKN2B* promoter and another one is located 130 kb upstream, in the CAD interval. Finally we searched for marks indicative of enhancers; enrichment of histone H3 monomethylation at lysine 4 (H3K4me1), binding of p300 and MED1, presence of DNase hypersensitivity sites (DHS) and depletion of H3K4me3<sup>16</sup>. Looking at these marks, we predict the locations of 33 enhancers, of which 26 are significantly enriched in conserved sequences ( $p < 0.01$  – Supplemental Table 2). Six enhancers are proximal to the CAD interval; nine enhancers are located in the CAD interval (referred to as ECAD1-9), two in the T2D interval (referred to as ET2D1-2) and 16 distal to the T2D interval. The majority of the 33 predicted enhancers fall in the proximal part of the gene-desert, in or near the CAD and T2D interval. These findings were further supported by the analysis of publicly available genome-wide datasets generated to predict regulatory elements using a variety of cell types (Figure 1). Additionally, we validated the enhancer activity of the ECAD2, ECAD4, ECAD5, ECAD7, ECAD8, ECAD9 and 1 ET2D elements using luciferase reporter assays in HeLa cells (data not shown). Interestingly, we determined that the 9p21 interval is the second densest gene-desert for predicted enhancers (7.5 per 100kb) in the human genome and the one containing the most disease-associated variants (10 SNPs – Supplemental Table 3). These data indicate that the 9p21 gene desert has an important regulatory function.

To identify the complete set of DNA variants in the 9p21 gene desert we sequenced a 196-kb interval (from *CDKN2B* promoter to 65 kb distal of the T2D interval) in 50 individuals of European descent and used the variants to refine the pattern of LD in the interval (Figure 2a). We identified 765 variants (31 indels and 734 SNPs) and used  $r^2$  correlation coefficient to identify variants in LD with any of the 8 CAD or 4 T2D associated SNPs. Of the 765 variants, 41 spanning 44 kb are in perfect LD with a CAD associated SNP, increasing to 131 variants spanning 111 kb when lowering  $r^2$  threshold to 0.5 (Figure 2b). In contrast to the

CAD interval, only 23 variants were in LD ( $r^2 \geq 0.5$ ) with at least one of the 4 T2D associated SNPs, spanning 11-kb. No variants were in LD ( $r^2 \geq 0.5$ ) with both CAD and T2D associated SNPs.

The functional variants underlying the association of 9p21 with CAD and T2D are highly likely a subset of the 131 and 23 candidate variants in LD ( $r^2 \geq 0.5$ ) with the initial CAD or T2D associated SNPs, respectively. However, the strong LD structure in the 9p21 gene-desert presents a hurdle for the precise identification of the causative functional variants using only genetic evidence. Thus, we determined which of the candidate variants are located in functional elements. Five of the CAD candidate variants are located in exons and 118 in introns of *CDKN2BAS*, however, none of the exonic variants are located in a conserved element suggesting that they are unlikely to be functional. Thirty-three CAD candidate variants are located in ECAD1-9 sequences; ECAD9 is significantly enriched for variants (11 in 1700 bp;  $p < 3.1 \times 10^{-5}$ ). The remaining 22 variants are distributed in ECAD1 (3 variants), ECAD2 (2), ECAD3 (8), ECAD4 (1), ECAD5 (3), ECAD6 (1), ECAD7 (3), and ECAD8 (1). Six of the 23 T2D candidate variants are located in a T2D enhancer, 5 of which are in ET2D1 and one in ET2D2.

To identify candidate regulatory variants, we computationally determined which variants disrupt consensus transcription factor-binding sites (TFBS) in the predicted enhancers (Supplemental Tables 4 & 5). Although many of the enhancers in the CAD interval had variants disrupting TFBS, ECAD9 containing 8 such variants was the most notable. Additionally, the SNP (rs10757278) most consistently associated with increased risk of CAD<sup>17</sup> is in ECAD9 and disrupts a TFBS involved in the inflammatory response (STAT1 – Supplemental Table 4). The rs10811656 and rs10757278 SNPs are located 4-bp apart in the predicted ECAD9 STAT1 binding site in the non-risk haplotype, which is disrupted in the risk haplotype (TTCCGGTAA > TTCTGGTAG). STAT1 has previously been shown to bind this locus in HeLa cells<sup>18</sup>, which are heterozygous for rs10811656/rs10757278 alleles, when treated with IFN $\gamma$ . We observed the binding of STAT1 at the same locus in HUVEC heterozygous for the rs10811656/rs10757278 alleles upon treatment with IFN $\gamma$  (Figure 3a). We examined the effect of IFN $\gamma$  treatment in heterozygous HeLa and HUVEC cells on the expression of *CDKN2B* and *CDKN2BAS*. Interestingly, we noticed that IFN $\gamma$  treatment leads to the repression and induction of *CDKN2B* and *CDKN2BAS*, respectively. This effect is greater in HUVEC where *CDKN2BAS* expression is induced 4-fold and *CDKN2B* transcription is repressed 2-fold (Figure 3b). These results are consistent with the epigenetic transcriptional repression of *CDKN2B* induced by ANRIL, the transcript encoded by *CDKN2BAS*<sup>19</sup>.

To validate the disruption of STAT1 occupancy at the rs10811656/rs10757278 risk alleles, we used LCL homozygous for the CAD risk haplotype. STAT1 is constitutively expressed at high levels in LCL<sup>20</sup>. We showed by chromatin immunoprecipitation (ChIP) that STAT1 binds at ECAD9 in 2 LCL homozygous for the CAD non-risk haplotype (2.7-fold enrichment), whereas it does not bind in 2 LCL homozygous for the risk haplotype (Figure 3c). To further establish a functional link between STAT1 occupancy in ECAD9 and expression of *CDKN2BAS*, we assessed the expression of *CDKN2BAS* in LCL upon *STAT1* siRNA mediated knock-down; the expression of *CDKN2BAS* was significantly up regulated (7-fold) in LCL homozygous for the CAD non-risk haplotype, but the effect was quite small in LCL homozygous for the CAD risk haplotype (Figure 3d). These results are consistent with the fact that STAT1 fails to bind ECAD9 in the CAD risk haplotype and thus is unlikely to participate in *CDKN2BAS* regulation. Interestingly, these results suggest that the effects of STAT1 binding at ECAD9 may have cell type-specific differences on the expression of *CDKN2BAS* between LCL (repression) and HUVEC (activation) (Supplemental Discussion).

To determine which genes are interacting with and potentially regulated by the ECAD9 enhancers we examined all genes (20 upstream and 1 downstream) located within 2Mb for interactions. As long-range enhancer interactions are often tissue specific<sup>21</sup> we chose to use HUVEC cells as a model system for vascular endothelium. We performed chromatin conformation capture (3C) using cross-linked HUVEC grown in standard conditions, followed by BamH1 digestion of the chromatin and diluted religation<sup>22</sup>. We analyzed the 3C using a novel approach derived from combining 3C with DNA selection and ligation<sup>23</sup> (method referred to as 3D-DSL) and examined the ligated acceptor and donor BamH1 sites via high throughput sequencing. The resolution of 3D-DSL is limited by the distribution of BamH1 sites, to compensate for this we selected 6 BamH1 sites flanking ECAD9 and other neighboring enhancers as acceptor sites, and 145 BamH1 donor sites distributed throughout the 2 Mb of sequence tested for interactions. We identified 9 donor sites interacting strongly with the enhancer acceptor sites (Figure 4a) of which 7 are sufficiently distant (>45 kb) from the acceptor sites to be recognized as specific (Supplemental Table 6). They occur in 4 areas: 1) *CDKN2A/B* locus, 2) *MTAP* gene, 3) downstream of *IFNA21* and 4) downstream of the T2D Associated Region (DOTAR), an interval of unknown function. Accounting for the resolution limitations due to the distribution of BamH1 sites, we did not assign ligated acceptor sites as belonging to ECAD9 but rather refer to them as “associated enhancers” and grouped the donor site interactions in *CDKN2A* and *CDKN2B* as “*CDKN2A/B*”. We confirmed the interactions between the associated enhancers and surrounding genes identified by the 3D-DSL approach using traditional methods. The interaction between the associated enhancers and *IFNA21*, which spans 946 kb, was assayed by fluorescence *in situ* hybridization (FISH). We examined whether or not treatment with IFN $\gamma$  would affect this long-range interaction and observed that the enhancers and *IFNA21* are in close proximity in 58% vs 37% of the chromosomes analyzed in HUVEC nuclei in the presence and absence of IFN $\gamma$ , respectively ( $p < 1.63 \times 10^{-25}$  Supplemental Figure 2 and Supplementary Table 7). These data indicate that this interaction occurs basally, but is significantly remodeled upon treatment with IFN $\gamma$ . We validated more closely interacting loci via site specific PCR on the religated 3C DNA and verified that the interaction between the associated enhancers and *CDKN2A/B* locus is strengthened in the presence of IFN $\gamma$  whereas the interaction with *MTAP* is reduced (Figure 4b).

In conclusion, our study shows that IFN $\gamma$  stimulation and STAT1 binding at ECAD9 play important roles in regulating the expression of *CDKN2B* and *CDKN2BAS*. Our findings are consistent with previous studies observing a correlation between CAD risk variants and *CDKN2A/B* expression in lymphocytes<sup>10</sup>, and reduced expression of *CDKN2A/B* in cardiac and other tissues in a mouse model with the orthologous CAD interval deleted<sup>8</sup>. We also demonstrate that the associated enhancers physically interact with the intervals encoding *CDKN2A/B*, *MTAP* and *IFNA21* in HUVEC and that these interactions are remodeled upon IFN $\gamma$  treatment, thus suggesting that the STAT1 binding site in ECAD9 is a key regulatory sequence. STAT1 is a downstream effector of the pathway that mediates response to inflammation, which is associated with angiogenesis<sup>24</sup> and atherosclerosis pathogenesis<sup>25</sup> in endothelial tissue. Thus, in endothelial tissues, the biological effects of the rs10811656 and rs10757278 risk alleles disrupting the ECAD9 STAT1 binding site might be augmented during activation of the inflammatory response. It is likely that the unique enhancer landscape of 9p21 is governed by higher order chromatin structure and thus involved in different temporal- or tissue- specific expressions of additional genes than those identified in our study<sup>26</sup>. Future studies assessing the effects of variants on the 9p21 chromatin landscape under different physiological conditions and cells types will undoubtedly provide additional insights into the association of this interval with CAD and T2D susceptibility.

## Methods Summary

### HUVEC, LCL and IFN $\gamma$ treatment

The 2 HUVEC cell lines used in the study were derived from male Caucasian donors genotyped as heterozygotes for the 9p21 CAD associated SNPs. The 4 EBV transformed lymphoblastoid cell lines (LCL) were selected for their genotypes (2 homozygous CAD risk or 2 homozygous CAD non-risk) using the HapMap data<sup>27</sup>. Experiments were performed within 2–4 passages by treating cells with 100 ng/ml of IFN $\gamma$  (R&D) for 24 hrs. Treated and untreated cells were subjected to RT-PCR (HUVEC), ChIP (HUVEC and LCL), 3C analysis (HUVEC) or FISH (HUVEC).

### Direct DNA Selection and Ligation (DSL) and sequencing of 3C interactions

**Probe design:** We designed 12 acceptor probes in the interval chr9:22100523-22126469 (hg18; spanning from ECAD7 to the ET2D2 enhancer), on both strands immediately 3' of the 6 BamHI sites in the region. We designed 290 donor probes on both strands immediately 5' of the 145 BamHI sites in the interval chr9:21035934-22494089 (hg18; except where acceptor probes were designed). A universal sequence added to the probes is compatible with Illumina GA adapters for direct sequencing. The 12 acceptor probes and 290 donor probes (Supplemental Table 8) were pooled in equimolar amounts, separately. **3D-DSL sequencing:** The DSL ligation products were prepared as described in Kwon et al.<sup>23</sup>. 3C was performed as per Lieberman-Aiden et al.<sup>28</sup> and the products were sheared by sonication. The sheared DNA (200–600 bp) was purified and biotinylated. Donor and acceptor probes pools were annealed to the biotinylated 3C samples and the biotinylated DNA was bound on to streptavidin magnetic beads. The 5' phosphate of acceptor probes and 3' OH<sup>-</sup> of donor probes were ligated using Taq DNA ligase. The ligated products were washed and eluted from the streptavidin magnetic beads, followed by PCR amplification and deep sequencing on the Illumina GA2 (Supplemental Information).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank X. Wang, K. Trevarthen and M. Nakano for experimental assistance. We thank the Scripps Genomic Medicine Clinical team for sample collection. This work was partially supported by NCCR grants 1U54RR025204 and 1UL1RR031980-01 and NIH grants HL065445, DK39949, DK018477, DK074868 and 1R21CA152613-01. MGR is an Investigator with HHMI.

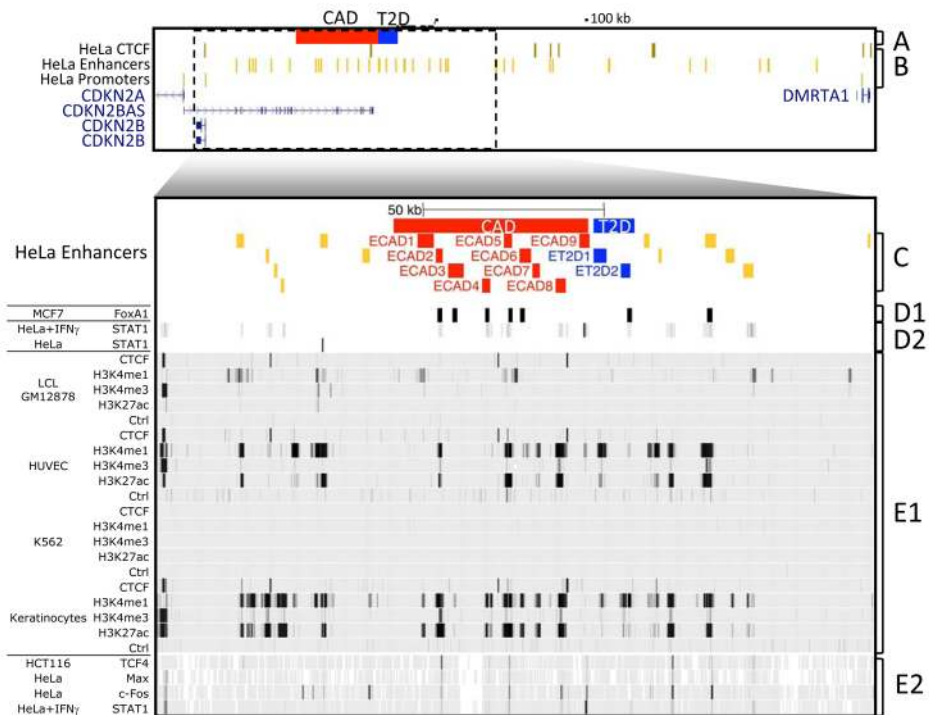
## References

1. McPherson R, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007; 316:1488–91. [PubMed: 17478681]
2. Helgadóttir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007; 316:1491–3. [PubMed: 17478679]
3. Samani NJ, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007; 357:443–53. [PubMed: 17634449]
4. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–78. [PubMed: 17554300]
5. Scott LJ, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007; 316:1341–5. [PubMed: 17463248]
6. Saxena R, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007; 316:1331–6. [PubMed: 17463246]



7. Zeggini E, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008; 40:638–45. [PubMed: 18372903]
8. Visel A, et al. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature.* 464:409–12. [PubMed: 20173736]
9. Jarinova O, et al. Functional Analysis of the Chromosome 9p21.3 Coronary Artery Disease Risk Locus. *Arterioscler Thromb Vasc Biol.* 2009
10. Liu Y, et al. INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS ONE.* 2009; 4:e5027. [PubMed: 19343170]
11. Helgadottir A, et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet.* 2008; 40:217–24. [PubMed: 18176561]
12. Newton-Cheh C, et al. A Common Variant at 9p21 Is Associated With Sudden and Arrhythmic Cardiac Death. *Circulation.* 2009 Published online on November 9, 2009.
13. Pasmant E, et al. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res.* 2007; 67:3963–9. [PubMed: 17440112]
14. Heintzman ND, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature.* 2009
15. Kim TH, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell.* 2007; 128:1231–45. [PubMed: 17382889]
16. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 2007; 39:311–8. [PubMed: 17277777]
17. Tousoulis D, Briasoulis A, Papageorgiou N, Antoniadis C, Stefanadis C. Candidate gene polymorphisms and the 9p21 locus in acute coronary syndromes. *Trends Mol Med.* 2008; 14:441–9. [PubMed: 18786860]
18. Robertson G, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods.* 2007; 4:651–7. [PubMed: 17558387]
19. Yap KL, et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell.* 38:662–74. [PubMed: 20541999]
20. Najjar I, et al. Latent membrane protein 1 regulates STAT1 through NF-kappaB-dependent interferon secretion in Epstein-Barr virus-immortalized B cells. *J Virol.* 2005; 79:4936–43. [PubMed: 15795279]
21. Ahmadiyeh N, et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci U S A.* 107:9742–6. [PubMed: 20453196]
22. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science.* 2002; 295:1306–11. [PubMed: 11847345]
23. Kwon YS, et al. Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters. *Proc Natl Acad Sci U S A.* 2007; 104:4852–7. [PubMed: 17360330]
24. Battle TE, Lynch RA, Frank DA. Signal transducer and activator of transcription 1 activation in endothelial cells is a negative regulator of angiogenesis. *Cancer Res.* 2006; 66:3649–57. [PubMed: 16585190]
25. Ross R. Atherosclerosis is an inflammatory disease. *Am Heart J.* 1999; 138:S419–20. [PubMed: 10539839]
26. Lomvardas S, et al. Interchromosomal interactions and olfactory receptor choice. *Cell.* 2006; 126:403–13. [PubMed: 16873069]
27. Frazer KA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007; 449:851–61. [PubMed: 17943122]
28. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009; 326:289–93. [PubMed: 19815776]

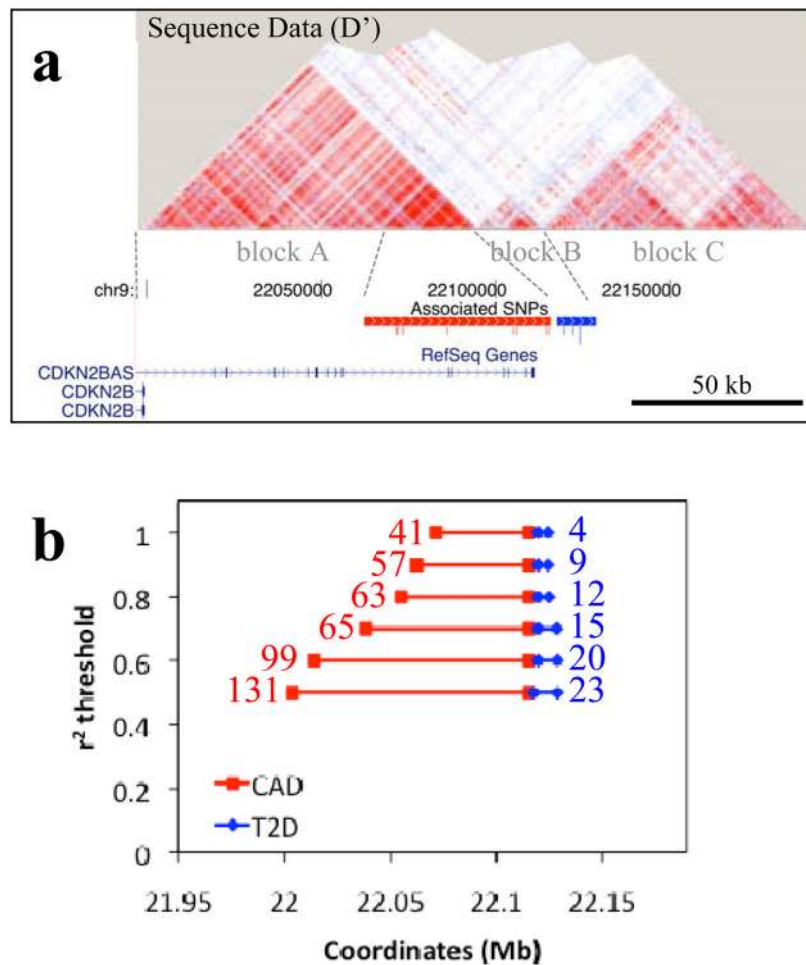
29. Lupien M, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*. 2008; 132:958–70. [PubMed: 18358809]
30. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*. 2004; 306:636–40. [PubMed: 15499007]



**Figure 1. Functional annotation of the 9p21 interval**

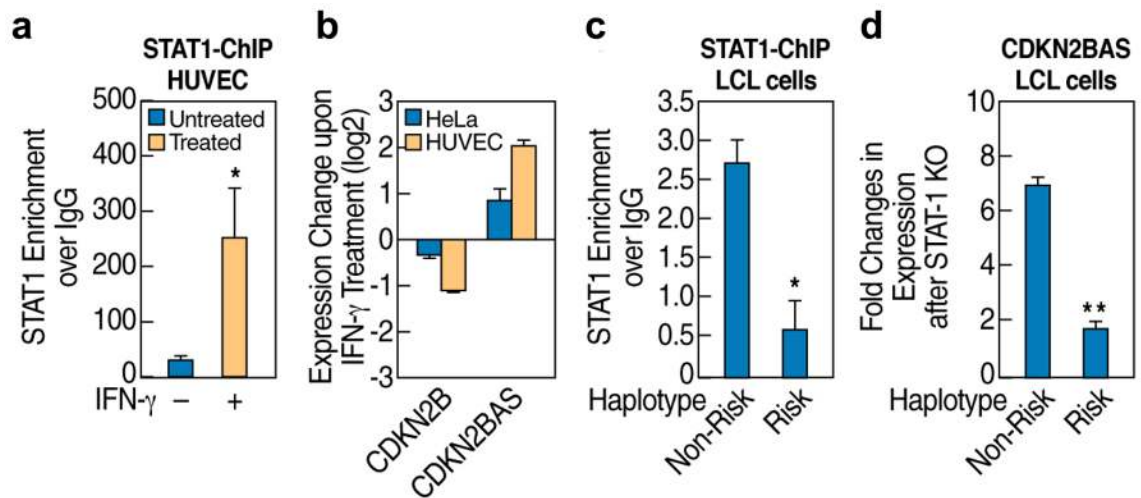
The locations of the core CAD and T2D associated intervals (track A) and the predicted insulators (brown), enhancers (orange) and promoters (green) in HeLa cells<sup>14</sup> are indicated (track B). The enhancers are distributed (track C) between the CAD interval (red), the T2D interval (blue) or located outside (orange). The location of the binding sites for FoxA1 in MCF7 cells (track D2)<sup>29</sup> and STAT1 in IFN $\gamma$  treated and non-treated HeLa (track D3)<sup>18</sup> as well as the distribution of 9p21 chromatin marks in the ENCODE data<sup>30</sup> (tracks E1 and E2 – Supplemental Methods) are indicated.





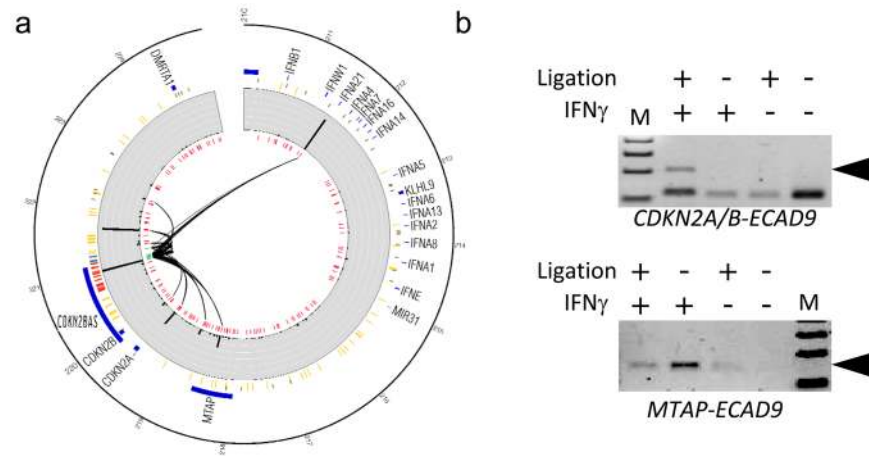
**Figure 2. LD analysis of the 9p21 interval**

(a) The sequenced interval shows three LD blocks. The 53-kb CAD interval is located at the 3' end of the first block (A), the second block (B) spans 11-kb corresponding to the T2D interval and the third block (C) spans 63-kb. LD map (D') based on variants identified in the 50 sequenced samples. (b) Number of variants in LD at various  $r^2$  thresholds (y axis) with any CAD (red) or T2D associated variants (blue). The distance spanned by the SNPs in LD is indicated (x axis).



**Figure 3. *In vivo* effects of the ECAD9 variants**

(a) Enrichment of the ECAD9 STAT1 binding site by anti-STAT1 ChIP in HUVEC cells untreated or treated with IFN $\gamma$ . (b) Changes in level of expression of *CDKN2B* and *CDKN2BAS* genes upon treatment with IFN $\gamma$  in HeLa and HUVEC. (c) Enrichment of the ECAD9 STAT1 binding site by anti-STAT1 ChIP in LCL homozygous for the CAD non-risk or CAD risk haplotypes. (d) Expression level changes of *CDKN2BAS* in LCL homozygous for non-risk or risk CAD haplotype after STAT1 knock-down by siRNA. (\*) and (\*\*) indicate two-tailed Student T-test  $p < 0.05$  and  $p < 0.01$ , respectively.



**Figure 4. Long-Range interaction with the enhancer locus**

(a) Circular representation of the 3D-DSL results. The scale shows chromosome 9 position (hg18) in 100kb increments. RefSeq genes (dark blue) and HeLa predicted enhancers (orange) are displayed with the CAD (red) and T2D (blue) associated enhancers. The histogram represents the normalized average number of reads mapping to each BamHI donor site. The inner circle links connect BamHI acceptor sites for the 9 most strongly interacting donor sites. (b) PCR validation of the long-range interaction between ECAD9 and *CDKN2A/B* (upper panel) or *MTAP* (lower panel). Arrow indicates the specific product.