MolecularMedicine

Three Interacting Genomic Loci Incorporating Two Novel Mutations Underlie the Evolution of Diet-Induced Diabetes

Yoram Yagil,[1,2] Barak Markus,[3] Refael Kohen,[3] and Chana Yagil[1,2]

[1]Laboratory for Molecular Medicine and Israeli Rat Genome Center, Barzilai University Medical Center, Ashkelon, Israel; [2]Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheba, Israel; and [3]The Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science, Rehovot, Israel

We investigated the pathophysiology of diet-induced diabetes in the Cohen diabetic rat (CDs/y) from its induction to its chronic phase, using a multilayered integrated genomic approach. We identified by linkage analysis two diabetes-related quantitative trait loci on RNO4 and RNO13. We determined their functional contribution to diabetes by chromosomal substitution, using congenic and consomic strains. To identify within these loci genes of relevance to diabetes, we sequenced the genome of CDs/y and compared it to 25 other rat strains. Within the RNO4 locus, we detected a novel high-impact deletion in the *Ndufa4* gene that was unique to CDs/y. Within the RNO13 locus, we found multiple SNPs and INDELs that were unique to CDs/y, but were unable to prioritize any of the genes. Genome-wide screening identified a third locus not detected by linkage analysis that consisted of a novel high-impact deletion on RNO11 that was unique to CDs/y and involved the *Sdf2l1* gene. Using cosegregation analysis, we investigated in silico the relative contribution to the diabetic phenotype and the interaction among the three genomic loci on RNO4, RNO11 and RNO13. We found that the RNO4 locus plays a major role during the induction of diabetes, whereas the genomic loci on RNO13 and RNO11, while interacting with the RNO4 locus, contribute more significantly to the diabetic phenotype during the chronic phase of the disease. The mechanisms whereby the mutations on RNO4 and -11 and the RNO13 locus contribute to the development of diabetes are under continuing investigation.

Online address: http://www.molmed.org
doi: 10.2119/molmed.2016.00114

INTRODUCTION

The pathophysiological basis of diabetes remains elusive, most likely because of the heterogeneity of the disease, the complexity of the pathogenesis involved and the simultaneous involvement of multiple pathways that act, interact and are regulated by a multitude of not yet fully understood variables. The complexity increases further when separate mechanisms are likely to come into play during the different stages in the development of the disease. The evolution of type 2 diabetes in humans usually includes sequential transition from health through a prediabetic state to frank diabetes, each phase evolving over a variable number of years. In experimental models of diabetes, the evolution is typically biphasic, incorporating an initiation phase followed by a chronic phase.

In the current study, we investigated the pathophysiology of diabetes in the Cohen rodent model of diet-induced diabetes (1). In this experimental model, diabetes is induced during the initiation phase by feeding genetically susceptible animals over 1 month with a custom-designed diabetogenic diet (DD) that is copper deficient and β-casein rich, the two major dietary culprits that we have identified as contributing to the development of the disease (2). Thereafter, diabetes in this model evolves into its chronic phase over the subsequent months, as long as the animals continue to be fed with DD (3,2).

We have previously focused our studies on the pathophysiology of diabetes during the induction phase of diabetes after the first 4 weeks of exposure to DD (3,2). In the current study, we adopted an integrative genomic approach to investigate the evolution of diabetes during its different phases, from its initiation into its chronic phase. We hypothesized that distinct mechanisms are involved in the different phases of the disease. Specifically, we sought to identify genes that account for the diabetic phenotype at different time points during the evolution of diabetes, anticipating that these would lead to the mechanisms involved.

Our study led to the discovery of three genomic loci, two of which are novel, that are highly likely to be involved in

Address correspondence to *Yoram Yagil, Laboratory for Molecular Medicine and Israeli Rat Genome Center, Barzilai University Medical Center, 2 Hahistadrut Street, Ashkelon 78278, Israel. Phone: +972-50-6819833; Fax: +972-8-6745824; E-mail: labmomed@bgu.ac.il.*
Submitted April 29, 2016; Accepted for Publication July 25, 2016; Published Online (www.molmed.org) July 26, 2016.

Feinstein Institute for Medical Research
Northwell Health™

the various phases of the development of diabetes. In two of these loci, we identified novel high-impact mutations that account, at least in part, for the diabetic phenotype. We established the relative contribution of each locus to the diabetic phenotype and the interactions among the loci at each phase of the evolution of diabetes in our model.

MATERIALS AND METHODS

Animals

We utilized the Cohen diabetic rat as our basic experimental model, which consists of two substrains, the Cohen diabetic-sensitive (CDs/y) strain and the Cohen diabetic-resistant (CDr/y strain (1). For purposes of genetic mapping, we used the diabetes-prone CDs/y strain. As contrasting strain, we used the non-diabetic SBN/y strain (4), as opposed to our previous studies, in which we used CDr/y as the contrasting strain (3,2), aiming thereby to diversify the genomic background. We obtained both strains from the colonies inbred at the Israeli Rat Genome Center (IRGC) of the Barzilai University Medical Center in Ashkelon, Israel (www.irgc.co.il) (1,4). The strains were designated with the suffix "/y" to separate the IRGC colonies from those originally inbred at the Hebrew University.

We housed animals in accordance with institutional regulations, principles of laboratory animal care (National Institutes of Health publication no. 85-23, revised 1985) and guidelines of the American Society of Physiology for the care and use of laboratory animals. We maintained climate-controlled conditions and light to regularly timed diurnal cycles.

To elicit the diabetic phenotype in CDs/y, we weaned animals at 1 month and provided them with standard rat chow and tap water *ad libitum*. At 6–7 wks, we switched to DD and copper-poor water. We custom-prepared DD, as previously described (1).

Genetic Mapping

We intercrossed 11 female SBN/y with 3 male CDs/y rats, mated F1 progenies brother-to-sister and produced an F2 cohort consisting of 166 male animals. For reference, we studied parental SBN/y and CDs/y.

Phenotype. We studied the metabolic phenotype of F2 and parental strains over 4 months, using as surrogate blood glucose level (BGL) during an oral glucose tolerance test (OGTT). We performed the OGTT as previously described (1,2) in the same animals repeatedly at 1, 2.5 and 4 months after initiation of DD. We calculated the area under the curve (AUC) of BGL during the OGTT and used AUC as a quantitative measure reflecting the diabetic status. We provide the data as mean ± standard error and compared between groups using the Student *t* test or one-way analysis of variance (ANOVA), as applicable.

Genotype. We extracted genomic DNA from the liver by salt precipitation, as previously described (4). We determined genotype by PCR amplification of genomic DNA with marker (microsatellite)–specific primers that were polymorphic between CDs/y and SBN/y (5). On each chromosome, we targeted markers spaced 10–20 cm apart. When we detected a genomic locus of interest, we increased the marker density. Microsatellite markers were custom synthesized (Genosys, Sigma) using primer sequences provided by the Rat Genome Database (RGD) (http://www.rgd.mcw.edu).

Linkage. For linkage analysis, we used the MultiQTL software package, version 2.6 (www.multiqtl.com), as previously described (6). In brief, we initially screened the entire genome for genetic linkage using single trait analysis (STA), analyzing each time period (months after initiating DD) separately. When we detected linkage at consecutive time points, we applied multienvironment analysis (MENVA).

Cosegregation. We performed cosegregation analysis at each quantitative trait locus (genomic locus) detected by linkage analysis, as previously described (3,2). In brief, we compared by ANOVA the phenotype in F2 between those with CDs/y and/or SBN/y alleles, which allowed determination of the effect of genotype on phenotype (sensitivity or resistance to diabetes).

Chromosomal Substitution

To determine the functional contribution of the RNO4 quantitative trait locus we detected by linkage analysis to the diabetic phenotype, we utilized a previously constructed consomic strain in which RNO4 had been introgressed in full from CDr/y over the background of CDs/y, the CDs/y.4$^{CDr/y}$ strain (7). We then created a new congenic strain, as previously described (4), in which we introgressed with 10 backcrosses the RNO4 locus from CDr/y onto the genomic background of CDs/y, generating the CDs/y.RNO-4Locus$^{CDr/y}$ strain (defined by microsatellite markers D4Rat9 and D4Rat153 and spanning positions 23,991,721 and 42,637,464, NCBI, Rnor 6.0). In both strains, we performed an OGTT after 1 and 2.5 months of DD.

To determine the contribution of the RNO13 locus to the diabetic phenotype, we constructed an additional congenic strain in which we introgressed the RNO13 genomic segment from SBN/y onto the background of CDs/y with 10 backcrosses, generating CDs/y. RNO-13Locus$^{SBN/y}$ (defined by microsatellite markers D13Rat85 and D13Mit4 and spanning positions 74,568,378 and 92,916,783). We performed an OGTT in that strain after 1 and 2 months of DD.

Candidate Genes

To identify genes of relevance to diabetes within the confines of the genomic loci (QTLs) detected by linkage, we scanned web-based databases, including the University of California Santa Cruz (UCSC) Genome Browser (https://genome.ucsc.edu/) and the RGD browser and generated lists of potential candidate genes within the respective loci. Aiming to find sequence variations that might highlight within these lists the culprit genes, we then sequenced the genomes of CDs/y and

CDr/y using next-generation sequencing (NGS). In brief, we performed DNA-seq PCR-free as previously described (8), with the following modifications: Two samples of ~1 ug of genomic rat liver DNA were sheared to a peak of 291 and 286 bp, using the Covaris S220 sonicator. We performed end repair in 80 ul reaction at 20°C for 30 min. Following cleanup with Agencourt AmPURE XP beads (Beckman Coulter) at a ratio of 0.75 × Beads/DNA volume, we added bases to both 3′ ends. We ligated adapters (Bio Scientific, cat. no. 514112) in a final concentration of 1.1 uM following SPRI beads cleanup at a ratio of 0.75 × beads/DNA volume. We prepared the samples without a PCR step. Libraries quantitation was in 50 single-read HigSeq runs. We pooled both libraries and ran 6 lanes of 100 bp paired END on the Illumina V3 HiSeq 2500 instrument. The resulting raw sequencing data from CDs/y and CDr/y have been submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov.sra/) under accession number SRP067304.

Once the genomic sequences became available, we screened for sequence variations unique to CDs/y, initially within the confines of the QTLs and subsequently genome-wide outside the confines of the QTLs. We used the Brown Norway rat (Rattus norvegicus; BN) Rat Genome Sequencing Consortium (RGSC) Genome Assembly v6.0 (https://genome.ucsc.edu/; https://www.hgsc.bcm.edu/rat-genome-project) as reference genome and compared genome-wide the sequence of CDs/y to 25 other rat strains (CDr/y, ACI/EurMcwi, F344/NCrl, BBDP/Wor, FHH/EurMcwi, FHL/EurMcwi, GK/Ox, LE/Stm, LEW/Crl, LEW/NCrlBR, LH/MavRrrc, LL/MavRrrc, LN/MavRrrc, MHS/Gib, MNS/Gib, SBH/y, SBN/y, SHR/NHsd, SHRSP/Gla, SS/Jr, SR/Jr, SS/JrHsdMcwi, WAG/Rij, WKY/NCrl, WKY/NHsd) (9). We mapped raw fastq files to the reference genome using BWA software (0.7.5a-r405) with standard parameters (10). We inferred single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs) using GATK version 3.3 (11,12). Briefly, for all samples except CDs/y and CDr/y, we marked the duplicated reads and then inferred SNPs and INDELs jointly on all samples using the gvcf mode of the haplotype caller tool. Our endpoint was identification of SNPs and INDELs that were unique to CDs/y, initially within the genomic loci of interest and subsequently genome-wide. To infer copy number variations (CNVs), we used SPEEDSEQ (13) and LUMPY (14), followed by in-house scripts for filtering and population analysis. We conducted a genome-wide inference of large deletions, using the following filtering algorithm: (1) we selected only homozygous deletions; (2) we filtered loci with high-density deletions as false positives; (3) we filtered deletions that were near (<1000 bp) unknown sequence stretches (poly N in reference). We based the annotation on SNPeff version 4.1a (15). To generate high-confidence calls, we used mean coverage below 10 reads within the deletion, at least one single event of soft clipped reads supporting the deletion and at most five samples sharing some portion of the deletion. For population analysis, we compared deletions among different samples by selecting a reference sample in which all detected deletions were defined as reference events, treating all other samples as controls. We calculated the percentage of overlap between the reference events and events found in each of the control samples.

Interaction Between Genomic Loci

We assessed *in silico* the individual contribution of each locus to the diabetic phenotype (AUC) by applying the principle of cosegregation between genotype and phenotype, the magnitude of the phenotype reflecting the level of contribution. We then tested with a similar approach whether the individual loci interacted with one another in determining the phenotype by testing whether combinations of genotypes affected the phenotype differently. We incorporated in this mode of analysis data from the current F2 cross between CDs/y and SBN/y (166 animals), as well as relevant data from two previous F2 crosses between CDs/y and CDr/y (233 animals) (3,2).

All supplementary materials are available online at www.molmed.org.

RESULTS

Metabolic Phenotypes

Parental strains. In Cds/y-DD, (Figure 1) fasting BGL rose from 115 ± 10 mg/dL at 1 month to 141 ± 12 mg/dL at 4 months, whereas in SBN/y it remained <100 mg/L at all time points. After glucose loading, BGL in CDs/y-DD rose to a peak at 60 min, gradually dropping thereafter, with peak levels steadily increasing from 1 to 4 months, whereas in SBN/y-DD, BGL rose to a peak at 30 min to lower values than in CDs/y and did not increase over time. Consequently, AUC in CDs/y-DD was higher than in SBN/y-DD at all time points. AUC ranged from nondiabetic values as in SBN/y-DD, through intermediate values, up to overtly diabetic values as in Cds/y-DD (Figure 2).

Genomic Analysis of the F2 Cross

Using single trait analysis, we detected consistent and significant linkage between AUC and two genomic loci, one on RNO4 and the other on RNO13, at 1, 2.5 and 4 months after initiating DD (Figure 3, upper panel). The logarithm of the odds (LOD) score for the RNO4 QTL was highest after 1 month of DD and gradually declined after 2.5 and 4 months. The LOD score for the RNO13 QTL, on the other hand, remained elevated and unchanged over time (Figure 3, lower panel).

Using MENVA that integrated AUC after 1, 2.5 and 4 months, we confirmed the presence of these two QTLs, with the following features:

RNO4 locus (Figure 4, upper panel); highly significant by permutation analysis ($p = 0.0000$), maximal LOD score at position 26.5 ± 8.2 cM (100% power of

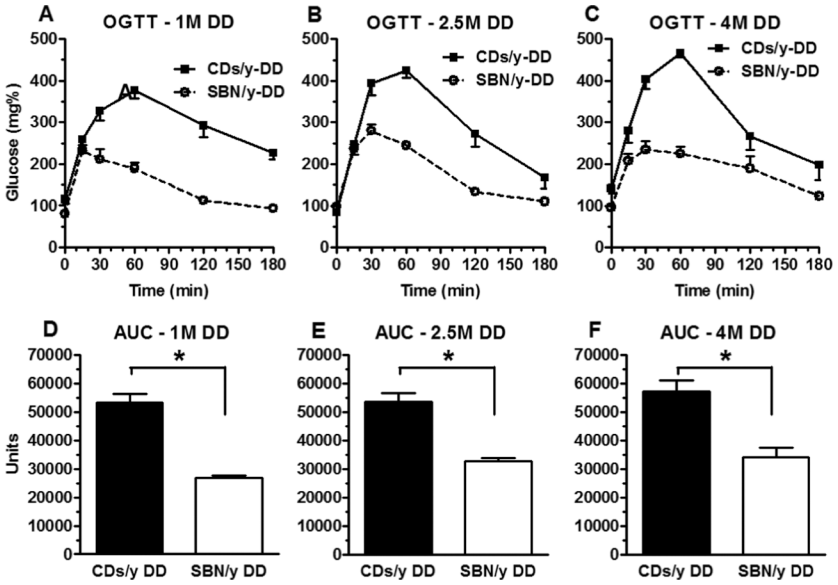

Figure 1. Results of the oral glucose tolerance test (OGTT) in CDs/y (n = 6) and SBN/y (n = 6) after 1, 2.5 and 4 months of feeding with diabetogenic diet (DD). The upper panels (A, B and C) represent average blood glucose levels at set times of the OGTT (mean ± standard error of the mean). The lower panels (D, E and F) represent the calculated area under the curve (AUC) of glucose levels during the OGTT by the respective time points. *$p < 0.01$ by unpaired Student *t* test.

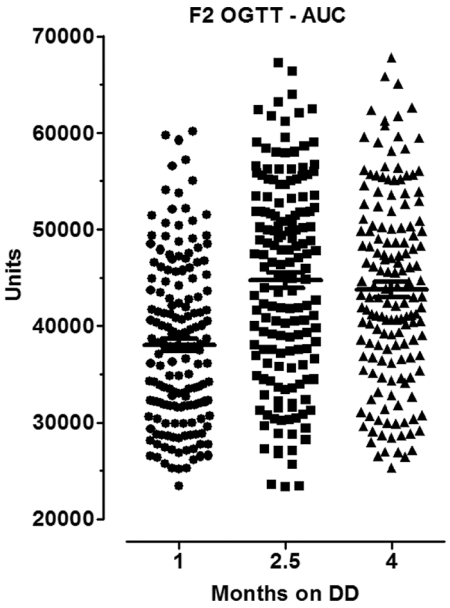

Figure 2. Results of the OGTT in F2 after 1, 2.5 and 4 months of feeding with DD. Data are presented as scattergram of the AUC of individual animals, as well as mean ± sem.

detection at the $p = 0.001$ level), span 16.1 cm, flanked by *D4Rat134* and *D4Rat151* (95% confidence interval [CI]) and contributing 17.9% to the phenotypic variation of the trait.

RNO13 locus (Figure 4, lower panel); highly significant by permutation ($p = 0.0000$), maximal LOD score at 33.1 ± 2.0 cm (100% power of detection at the $p = 0.001$ level), span 4.1 cM, flanked by *D13Rat57* and *D13Rat46* (95% CI) and contributing 16.5% to the phenotypic variation of the trait.

We confirmed these findings by cosegregation analysis, which demonstrated that at these two loci, the SS genotype was associated with a significantly higher AUC than in RR or SR (data not shown).

Chromosomal Substitution

RNO4 locus. In both consomic and congenic strains in which the entire chromosome or the QTL only from CDs/y was replaced with that of CDr/y, BGL during the OGTT and AUC after 1 month of DD was lower than in the diabetic CDs/y but higher than in the nondiabetic CDr/y (Figure 6, panels A and B). After 2.5 months of DD, however, BGL in the consomic and congenic strains were no longer different from CDs/y, yet were significantly higher than in CDr/y (Figure 6, panels C and D). These findings suggest that during the initiation phase of diabetes, the full expression of the diabetic phenotype is dependent upon the presence of CDs/y alleles at the RNO4 genomic locus, but much less so in the chronic phase of the disease, during which other genomic loci are likely to contribute more importantly to the phenotype.

RNO13 locus. In the congenic strain in which the QTL from CDs/y was replaced with that of CDr/y, BGL and AUC after 1 month of DD were not different from those in the diabetic CDs/y strain and were higher than in SBN/y and CDr/y (Figure 5, panels E and F). After 2 months of DD, the pattern remained unchanged (Figure 5, panels G and H).

Candidate Genes

To identify candidate genes within the RNO4 QTL, we used the narrow genomic span we previously reported in F2 crosses between CDs/y and CDr/y (3,2). Using the RGSC Genome Assembly v6.0, we detected within the span 33,637, 803–39,073,276 bp 15 protein-coding genes and 13 pseudogenes (Supplementary Table S1). Among these, we were able to single out *Ndufa4*, in which we found a novel, previously unreported large deletion with high impact involving exon 3 and unique to CDs/y (Figure 6A). We confirmed this deletion by demonstrating on agarose gel a difference in genomic DNA between CDs/y <1000 bp and CDr/y>3000 bp (Figure 6B), and by Sanger sequencing that revealed in CDs/y but not in CDr/y a deletion of 2793 bp from the end of intron 2 to the middle of intron 3 (Supplementary Figure S1). In addition to this major finding, we detected SNPs and/or INDELs in five other genes, one of which was unique to CDs/y and another to CDr/y, but all were within intronic regions and none within exons.

To identify candidate genes with the RNO13 QTL, we used the genomic span

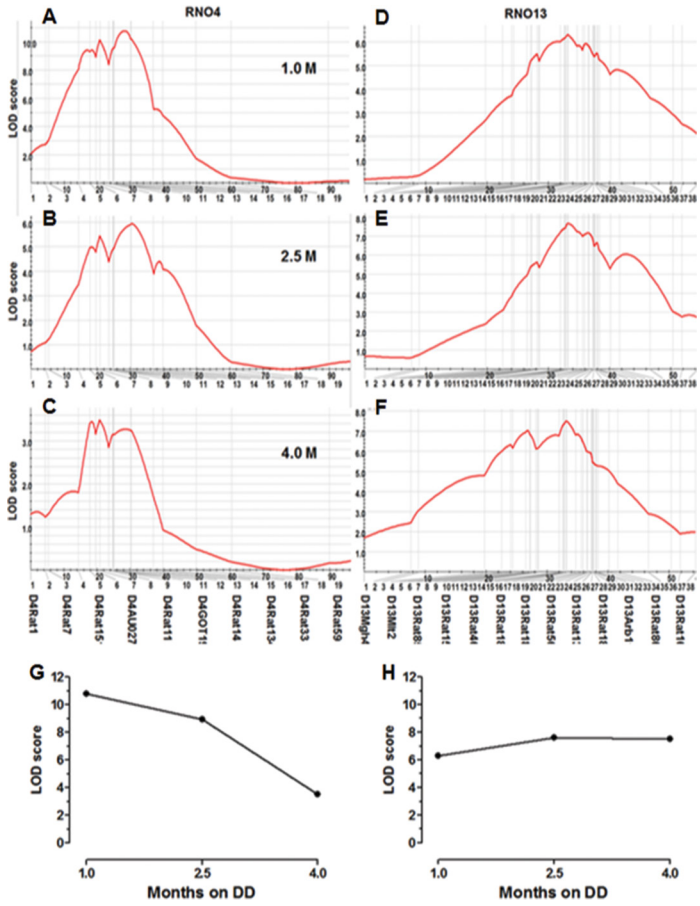

Figure 3. Logarithm of the odds (LOD) score of linkage analysis for OGTT AUC at 180 min on RNO4 (panels A, B and C) and RNO13 (panels D, E and F) after 1, 2.5 and 4 months of feeding with DD. Maximal LOD score of linkage analysis for OGTT AUC at 180 min loci on RNO4 (panel G) and RNO13 (panel H) after 1, 2.5 and 4 months of feeding with DD.

79,567,08–86,833,540 bp and identified 66 protein-coding genes (Supplementary Table S2) and 21 noncoding RNA sequences, three of which were Mirs (Mir1992a, Mir214 and Mir3120) and 19 pseudogenes. Of the 66 protein-coding genes, seven have been functionally or structurally associated with the endoplasmic reticulum (*Fmo6*, *Fmo13*, *Suco*, *Fmo4*, *F5*, *Fmo9* and *Tmco1*), five with the Golgi apparatus (*Vamp4*, *Scyl3*, *F5*, *Blzf1* and *Gorab*), and two with mitochondria (*Mpc2* and *Aldh9a1*). We identified a variable number of SNPs and/or INDELs within 41/66 genes, as well as in the Prrc2c pseudogene and in Mir214. In 26 genes, one or more SNP or INDEL were unique to CDs/y or CDr/y, but for lack of additional exceptional findings, we were unable to single out any of these genes.

We then screened the entire genome for additional high-impact gene variants not necessarily within the confines of the two QTLs on RNO4 and -13. We could not find any SNPs or INDELs that were unique to CDs or predicted to have a drastic impact on the function of the gene (frameshift, stop gain, stop loss and splice-site mutations). We did, however, detect a novel, previously unreported deletion on RNO11, which coincided with the *Sdf2l1* gene (NM_001109433, chr11: 88,122,327-88,123,234), incorporating two out of its three exons. Interestingly, this deletion was shared with the nondiabetic SBN/y strain but, importantly, none of the other 24 strains (Figure 6C). We confirmed this deletion by demonstrating on agarose gel a ~900 bp difference in genomic DNA between CDs/y and CDr/y (Figure 6D) and by Sanger sequencing that revealed a deletion of 907 bp in CDs/y and SBN/y, but not in CDr/y (Supplementary Figure S2).

Interaction Among Genomic Loci During the Induction of Diabetes

To determine the relative contribution of each locus independently (single locus effect) to the diabetic phenotype (AUC) during the initiation phase of diabetes, and the level of interaction between the three loci (two- and three-loci effect), we used two F2 populations (combined) from previous intercrosses between CDs/y and CDr/y (3,2) and tested for cosegregation between the AUC and the genotype of microsatellite markers representative of the QTLs. The results are shown in Table 1.

Single locus effect. After 1 month of DD, F2 animals that were homozygous for CDs/y (SS genotype) at each of the three genomic loci (on RNO4, -11 and -13) had a higher AUC than heterozygotes (SR) or homozygous for the nondiabetic CDr/y (RR). The AUC, however, was lower than in parental CDs/y, suggesting that each locus alone in its homozygous SS form significantly impacts the diabetic phenotype, but does not by itself account for the full expression of diabetic phenotype.

Two loci effect. In studying the interaction between RNO4 and -11, we found that F2 homozygous with SS genotype at both loci had a higher AUC than when one of the loci was SR or RR. Similar results were obtained for RON4 and -13. These findings suggest interactions between the homozygous forms of RNO4 and -11 and of RNO4 and -13 that augment the expression of the diabetic phenotype during the initiation phase of diabetes. In studying the interaction between RNO11 and -13, we found that the genotype at either locus did not significantly affect the diabetic phenotype, as

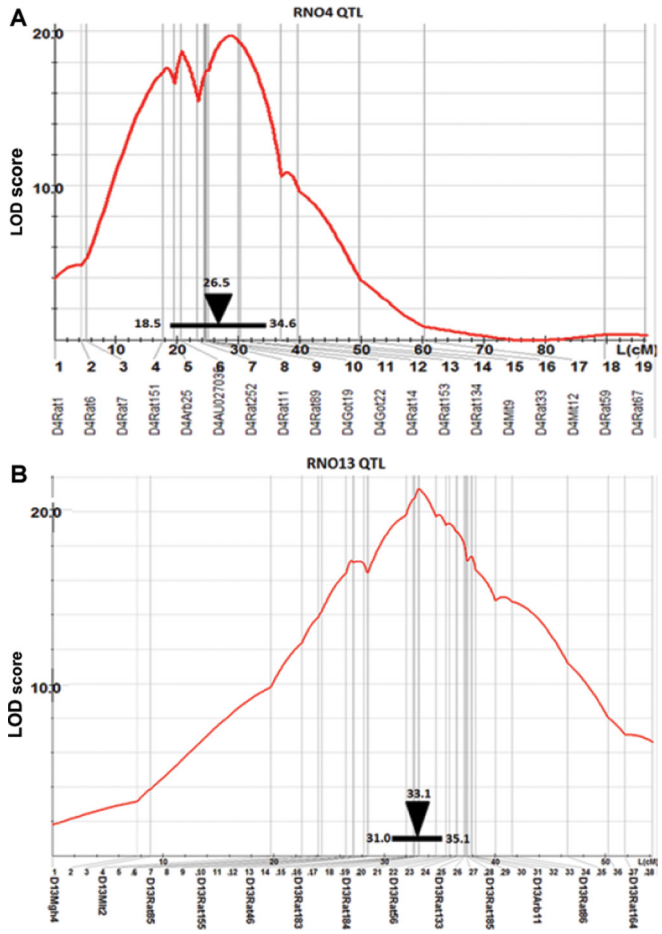

Figure 4. Results of linkage analysis, using the multienvironment mode of analysis and incorporating the data after 1, 2.5 and 4 months of DD. Data are expressed as LOD score tracing for the AUC of the OGTT over 180 min on RNO4 (panel A) and RNO13 (panel B). The horizontal bar represents the quantitative trait locus 95% confidence interval; the inverted triangle represents the point of the maximal LOD score.

there was no difference in the AUC if the genotype was SS at both loci or at only one locus and SR or RR at the other, suggesting a lack of additive effect between these two loci alone.

Three Loci Effect

When the genotype of F2 at all three genomic loci (RNO4, -11 and -13) was either SR or RR, the AUC was the lowest and not different from that found in parental CDr/y. In contrast, when the genotype at all three loci was SS, albeit the small number, the phenotype was maximal. When the RON4 locus and one more locus (whether RNO11 or RNO13) was SS, the diabetic phenotype was markedly enhanced, whereas when the RNO4 locus was SR/RR with loss of one or both S alleles, the impact on the AUC phenotype of the two other loci was reduced. These results highlight the major impact of the RNO4 locus on the diabetic phenotype during the initiation phase of diabetes.

Interaction Among the Genomic Loci During Progression of Diabetes

We studied the interaction among the three genomic loci after 1, 2.5 and 4 months of DD, using the F2 population from the current intercross between CDs/y and SBN/y. The results are provided in Table 2. When the genotype at all loci (RNO4, -11 and -13) was SS, the AUC was the highest after 1, 2.5 and 4 months, demonstrating an additive interaction among the three genomic loci. In the absence of one or two S alleles from the RNO13 locus and in the presence of two S alleles on the RNO4 and RNO11 loci (equivalent to the RNO13 congenic), the phenotype after 1 month of DD remained diabetic, confirming similar data derived from the CDs/y-CDr/y intercross and the results obtained in the RNO13 congenic strain after 1 and 2.5 months of DD. Importantly, after 4 months of DD, the diabetic phenotype diminished in the absence of one or two S alleles from the RNO13 locus despite the presence of homozygous SS loci on RNO4 and -11, suggesting an increasing role of the RNO13 locus and a diminished contribution of the RNO4 locus during the chronic phase of diabetes. In the presence of both SS alleles in the RNO11 and -13 loci and the absence of one or both SS alleles from the RNO4 locus, the diabetic phenotype 1 month after initiating DD was significantly attenuated, while at 2.5 and 4 months, the diabetic phenotype in F2 animals with the same genotype was no longer attenuated and the AUC was not different from the animals with both SS alleles in all three loci. These findings further support the impact of the RNO4 locus primarily during the induction phase of diabetes and its lesser contribution to the diabetic phenotype thereafter.

DISCUSSION

In the current study, we detected for the first time three distinct genomic loci, a genomic triad, that participate in the pathophysiology of diabetes in the Cohen diabetic rat model. Each of these loci contributes independently to the development of diabetes, but also interact with one another, allowing us to define their relative contributions to the metabolic phenotype. The locus on RNO4 was a confirmation of our previous findings, but the loci on RNO11 and -13 were entirely novel. Within two of these loci on RNO4 and -11, we identified novel

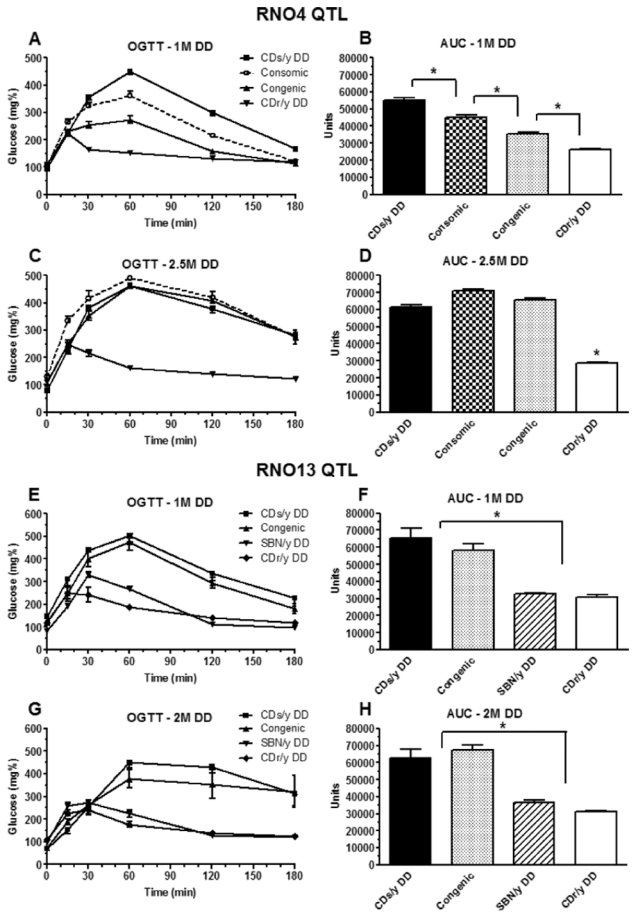

Figure 5. Results of the OGTT after 1 and 2.5 or 2 months of DD in consomic/congenic strains and their respective controls. The left panels (A and C) represent average blood glucose levels at set times of the OGTT (mean ± standard error of the mean). The right panels (B, C, F and G) represent the AUC of glucose levels during the OGTT. *$p < 0.01$ by ANOVA and post hoc LSD testing. RNO4 QTL: CDs/y (n = 29, 19), consomic strains for RNO4 (n = 29, 7), congenic strains for the RNO4 QTL (n = 25, 5) and CDr/y (n = 10, 4). RNO13 QTL: CDs/y (n = 8,6), congenic strains for the RNO13 QTL (n = 8,7), SBN/y (n = 9,10) and CDr/y (n = 8,5).

high-impact gene mutations that render them highly likely candidate genes for the pathophysiology of diabetes.

In our previous studies, focusing primarily on the initiation phase of diabetes, we detected after 1 month of DD one major genomic locus on RNO4, which failed, however, to explain the full genetic variation of the diabetic phenotype at that time point (3,2). In the current study, we confirmed the same genomic locus on RNO4 but established, based on a high LOD score at 1 month that gradually diminished after 2.5 and 4 months, that its contribution to the diabetic phenotype was major only during the induction of diabetes and gradually diminished thereafter. To support these findings, we resorted to the chromosomal substitution strategy. In the consomic strain introgressing the entire chromosome from CDr/y onto CDs/y and a congenic strain introgressing the locus only from CDr/y onto the genetic background of CDs/y, the diabetic phenotype was significantly attenuated after only 1 month of DD, but no longer after 2.5 months. These finding are consistent with our interpretation that the RNO4 locus plays a major role in the initiation phase of diabetes, and that this role diminishes thereafter.

We uncovered in the current study a novel second genomic locus on RNO13 already during the induction phase of diabetes after 1 month of DD. In contrast to RNO4, however, its LOD score remained elevated, highly significant and largely unchanged throughout the remaining study period at 2.5 and 4 months of DD. These results suggest that genes within the RNO13 locus are likely to be involved in the pathogenesis of diabetes during both induction of the disease and its maintenance phase, as opposed to the role of the RNO4 locus, mostly during the initiation phase. To support the findings relating to the role of the RNO13 QTL, we again resorted to the chromosomal substitution strategy. The results of the studies with the RNO13 congenic, however, failed to provide the direct evidence we sought to demonstrate that the RNO13 locus contributes independently to the expression of the diabetic phenotype during the induction or chronic phase of diabetes. We estimated the lack of confirmation as a false negative finding, for several possible reasons. The likelihood that we failed to capture the RNO13 QTL while generating our congenic strain is small, as the introgressed genomic segment from SBN/y to CDs/y was 3–4 cm wider than the 95% CI segment defined for the QTL. There are other possibilities that could account for the failure of the single congenic to exert its predicted effect, including that in order to contribute to the development of diabetes, the RNO13 QTL needs to interact with one or more QTLs. This latter explanation is the most likely, as demonstrated in the three-level *in silico* analysis that showed that for the diabetic phenotype to be fully expressed, RNO13 had to interact either with the RNO4 QTL alone or with both RNO4 and RNO11 QTLs. It is thus likely that the contribution of the QTL on RNO13 alone to the diabetic phenotype is only partial and not detectable using the chromosomal substitution strategy.

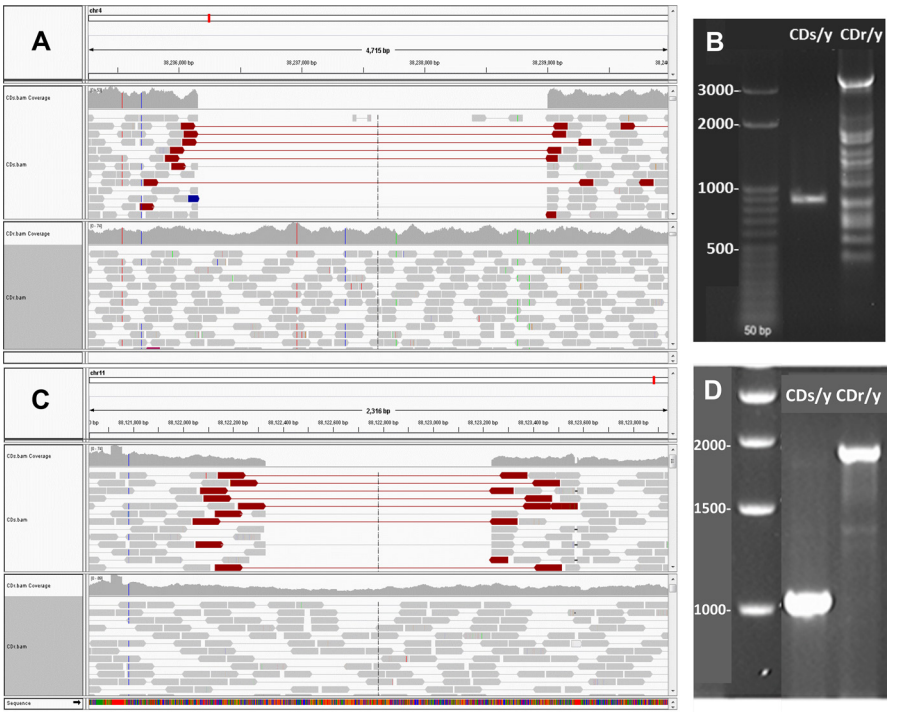

Figure 6. Large rare deletions found on RNO4 (panel A) and RNO11 (panel C). Each panel shows the reads and coverage tracks that allowed detection of large rare deletions in CDs/y but not in CDr/y in both chromosomes. Panels B and D demonstrate the difference in genomic DNA between the size of CDs/y and CDr/y on RNO4 and RNO11, respectively, on agarose gels.

The findings relating to the contribution of the two QTLs on RNO4 and -13 to the evolution and development of the diabetic phenotype lend support to our underlying hypothesis that different mechanisms contribute to the pathophysiology of diabetes during its various phases of evolution. Such findings further suggest that one set of genes appears to be involved primarily in the initiation phase, whereas other sets of genes contribute equally throughout the evolution of the disease, from its onset to its chronic phases.

Detection of QTLs of relevance to diabetes was, however, only the first step in the current study. To identify within the span of the QTLs genes of relevance to the evolving diabetic phenotype, we initially screened multiple published databases. Within the span of the RNO4 locus, we identified 15 protein-coding genes, but within the RNO13 locus, we identified a larger number of candidate genes, including 66 protein-coding genes. To select from these lists of candidate genes those most likely to be related to diabetes, we adopted the strategy of scanning the genome for sequence variations that were unique to the diabetes-prone CDs/y strain (susceptibility to diabetes). Our working assumption in adopting this strategy was that identifying unique sequence variations would lead us to the identities of genes likely to be related to the pathophysiology of diabetes. To apply this strategy, we sequenced the whole genome of CDs/y and CDr/y using the NGS platform and compared the genomic sequence of CDs/y to those of 25 other rat strains. We initially focused on the genomic segments identified by linkage, and then sought sequence variations genome-wide.

Within the span of the locus on RNO4, we detected in *Ndufa4* a novel, previously unreported, high-impact deletion that was unique to CDs/y. We had previously identified *Ndufa4* as a major candidate gene for diabetes, but we had failed at that time to detect within that gene the mutation that we currently found (3). Our current findings thus single out *Ndufa4* as the most likely candidate gene within that RNO4 QTL. *Ndufa4* has an important role in mitochondrial function, but its exact role in the pathophysiology of diabetes remains to be defined.

Within the span of the RNO13 QTL, we identified a large number of sequence variations in two-thirds of the genes within the locus, and in ~40% of the listed genes, one or more sequence variations were unique to CDs/y, some of them associated with endoplasmic reticulum, Golgi apparatus or mitochondriae. We were unable, however, to prioritize among the 26 genes that had unique sequence variations within the RNO13 locus.

The genome-wide screen for sequence variations led us to detect one additional novel, large-sequence variation on RNO11, consisting of a previously unreported large deletion within the *Sdfl2l1* gene in CDs/y. The *Sdfl2l1* gene has a role in the degradation of pro-insulin in the pancreatic β cells (16). The SDF2L1 protein contains 2 MIR domains and is similar to a family of SDF2 proteins that function when excess demand of unfolded proteins accumulates due to stress in the endoplasmic reticulum, for example, in diabetes and cancer (17). A role of the mutated *Sdfl2l1* gene in the pathophysiology of diabetes in our model is thus plausible, but remains to be better defined. Interestingly, this RNO11 mutation also appeared in SBN/y, which is a nondiabetic strain. However, a review of previous data in which an OGTT had been performed in SBN/y (data not shown) reveals that the AUC in SBN/y is usually mildly above that of CDr/y, which lacks the mutation, but significantly lower than CDs/y that carries at least one additional mutation within the RNO4 locus and a diabetogenic locus on RNO13. These data suggest that the RNO11 locus alone contributes to the diabetic phenotype, but only very mildly when on the SBN/y genomic background.

Table 1. Cosegregation analysis for representative microsatellite markers within the genomic loci on RNO4, -11 and -13.

				AUC			ANOVA		LSD test
			Genotype	AVG	SEM	N	F	P	SS vs RR/SR
Single locus effect		RNO4 locus	SS	40366	1969	44	46.02	<0.0001	P < 0.0001
			SR/RR	30511	533	188			
		RNO11 locus	SS	35896	1455	60	11.64	0.0006	P = 0.0008
			SR/RR	31154	646	172			
		RNO13 locus	SS	35651	1506	52	9.38	0.0025	P = 0.0025
			SR/RR	31256	639	171			
Two loci effect		RNO4	RNO11	AVG	SEM	N	F	P	P value
	1	SS	SS	47996	2962	12			1 vs 2 0.0054
	2	SS	SR/RR	37505	2293	32	9.52	0.0002	1 vs 3 <0.0001
	3	SR/RR	SS	32871	1355	32			2 vs 3 0.6510
		RNO4	RNO13	AVG	SEM	N	F	P	P value
	1	SS	SS	48558	8151	6			1 vs 2 0.0278
	2	SS	SR/RR	38408	1787	37	6.09	0.0034	1 vs 3 0.0016
	3	SR/RR	SS	33967	1181	46			2 vs 3 0.0543
		RNO11	RNO13	AVG	SEM	N	F	P	P value
	1	SS	SS	39222	3451	13			
	2	SS	SR/RR	34976	1658	47	0.95	0.3902	—-
	3	SR/RR	SS	34461	1726	39			
Three loci effect	RNO4	RNO11	RNO13	AVG	SEM	N	F	P	P value
	SR/RR	SR/RR	SR/RR	28592	531	105			
1	SS	SR/RR	SR/RR	35135	1757	27			1 vs 2 0.0072
2	SS	SS	SR/RR	47244	3359	10	5.03	0.0115	1 vs 3 0.0628
3	SS	SR/RR	SS	46959	12431	4			2 vs 3 0.9669
1	SR/RR	SS	SR/RR	31661	15044	37			1 vs 2 0.1097
2	SR/RR	SS	SS	36944	2850	11	10.87	0.0001	1 vs 3 <0.0001
3	SS	SS	SR/RR	47244	3359	10			2 vs 3 0.0158
1	SR/RR	SR/RR	SS	33032	1252	35			1 vs 2 0.2592
2	SR/RR	SS	SS	36944	2850	11	3.82	0.0291	1 vs 3 0.0106
3	SS	SR/RR	SS	46959	1243	4			2 vs 3 0.0900
	SS	SS	SS	51754	9959	2			

Data are presented for single locus, two loci and three loci effects, as the mean (AVG) of the area under the curve (AUC) after 1 month of DD, grouped by genotypes RR, SS and SR. Group averages were compared by one-way ANOVA, providing F and *p* values; post hoc analysis was performed using the least significant difference (LSD) test. SS, homozygous for CDs/y; SR, heterozygous homozygous for CDr/y.

Having identified three distinct genomic loci on RNO4, -11 and -13, we sought to determine their absolute and relative contributions to the development of the diabetic phenotype and whether they interacted with one another. We resorted to a three-level *in silico* analysis, studying the impact of each locus alone, the interaction between two genomic loci and the interaction between all three loci. Even though the number of animals used in the cross was relatively large, dividing the F2 animals by genotype during the analysis created multiple groups with a progressively decreasing number of animals in each group and risking a type 2 error. We were able, nonetheless, to demonstrate the major role that the RNO4 genomic locus fills in the pathophysiology of the induction phase of diabetes, and the increasingly important role of the RNO13 locus during the chronic phase. In addition, we were able to show that the three genomic loci we detected on RNO4, -11 and -13 interacted with one another and accounted in full for the development of diabetes in our models. It is highly likely, therefore, that the mutations we detected in the genes in the RNO4 and -11 loci, along with the RNO13 genomic locus, account in full for the pathophysiology of diabetes in this model, during both the induction and chronic phases of the disease.

CONCLUSION

Based on our findings, we currently hypothesize that diabetes develops in the susceptible CDs/y strain through the action of a triad of interacting genomic loci on RNO4, -11 and -13, which incorporate all the genes required for diabetes to develop in the diabetes-prone animal fed DD. *Ndufa4* on RNO4 is likely to be a major player in the induction of diabetes. One or more genes on RNO13 and *Sdfl2l1* on RNO11 are likely to contribute more importantly to the pathophysiology of the chronic phase of the disease. In identifying the culprit genomic triad, we have taken an important step forward in elucidating the pathophysiological basis of diabetes. Our next step is to investigate in depth the respective candidate genes within each of the three genomic loci and the mechanisms by which they interact with dietary components, copper and casein in particular, in causing diabetes, aiming eventually to prevent the development of and/or cure the disease.

ACKNOWLEDGMENTS

The authors wish to acknowledge the contribution of Sima Benjamin and Shlomit Gilad from the Weizmann Institute of Science, Rehovot, Israel, for the DNA library preparation using a customized, PCR-free, in-house protocol.

DISCLOSURE

The authors declare that they have no competing interests as defined by *Molecular Medicine*, or other interests that might be perceived to influence the results and discussion reported in this paper.

Table 2. Cosegregation analysis for representative microsatellite markers within the genomic loci on RNO4, -11 and -13, based on F2 intercross between CDs/y and SBN/y.

				AUC								
	Genomic Locus			1 month DD			2.5 months DD			4 months DD		
	RNO4	RNO11	RNO13	AVG	SEM	N	AVG	SEM	N	AVG	SEM	N
1	NN	SS	NN	30458	819	18	35934	2081	18	38328	1874	18
2	SN	SS	SN/NN									
	NN	SS	SN	35057	830	74	40972	1095	72	40387	1190	70
3	SN/NN	SS	SS	39909	1316	30	49524	1492	30	48413	1444	30
4	SS	SS	SN/NN	44267	1303	28	50745	1637	28	46502	1143	25
5	SS	SS	SS	46918	2125	14	53861	2070	13	53854	2114	13
	ANOVA		F	21.01			16.75			11.30		
			P	0.0000			0.0000			0.0000		
	Post hoc		1 vs 2	0.0120			0.0315			0.3767		
	LSD test		1 vs 3	0.0000			0.0000			0.0002		
			1 vs 4	0.0000			0.0000			0.0031		
			1 vs 5	0.0000			0.0000			0.0000		
			2 vs 3	0.0014			0.0000			0.0000		
			2 vs 4	0.0000			0.0000			0.0033		
			2 vs 5	0.0000			0.0000			0.0000		
			3 vs 4	0.0172			0.5986			0.4230		
			3 vs 5	0.0020			0.1400			0.0641		
			4 vs 5	0.2413			0.2932			0.0155		

Data are presented for three loci interaction, as the mean (AVG) of the area under the curve (AUC) after 1, 2.5 and 4 months of DD, grouped by genotypes RR, SN and NN. Group average were compared by one-way ANOVA, providing F and *p* values; post hoc analysis was performed using the least significant difference (LSD) test. SS, homozygous for CDs/y; SN, heterozygous; NN, homozygous for SBN/y.

REFERENCES

1. Weksler-Zangen S, Yagil C, Zangen DH, Ornoy A, Jacob HJ, *et al.* (2001) The newly inbred cohen diabetic rat: a nonobese normolipidemic genetic model of diet-induced type 2 diabetes expressing sex differences. *Diabetes*. 50:2521–9.
2. Yagil C, Barkalifa R, Sapojnikov M, Wechsler A, Ben-Dor D, *et al.* (2007) Metabolic and genomic dissection of diabetes in the Cohen rat. *Physiol. Genomics*. 29:181–92.
3. Barkalifa R, Yagil Y, Yagil C. (2010) Sex-specific genetic dissection of diabetes in a rodent model identifies Ica1 and Ndufa4 as major candidate genes. *Physiol. Genomics*. 42:445–55.
4. Yagil C, Katni G, Rubattu S, Stolpe C, Kreutz R, *et al.* (1996) Development, genotype and phenotype of a new colony of the Sabra hypertension prone (SBH/y) and resistant (SBN/y) rat model of slat sensitivity and resistance. *J. Hypertens.* 14:1175–82.
5. Yagil C, Sapojnikov M, Katni G, Ilan Z, Zangen SW, *et al.* (2002) Proteinuria and glomerulosclerosis in the Sabra genetic rat model of salt susceptibility. *Physiol. Genomics*. 9:167–78.
6. Yagil Y, Roif D, Sapojnikov M, Ben-Dor D, Tobar A, *et al.* (2014) Unmasking of proteinuria in the course of genetic dissection of nonproteinuric diabetic nephropathy. *Physiol. Genomics*. 46:29–38.
7. Yagil Y, Hessner M, Schulz H, Gosele C, Lebedev L, *et al.* (2010) Geno-transcriptomic dissection of proteinuria in the uninephrectomized rat uncovers a molecular complexity with sexual dimorphism. *Physiol. Genomics*. 42A:301–16.
8. Blecher-Gonen R, Barnett-Itzhaki Z, Jaitin D, Amann-Zalcenstein D, Lara-Astiaso D, *et al.* (2013) High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* 8:539–54.
9. Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, *et al.* (2013) Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell*. 154:691–703.
10. Li H, Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 26:589–95.
11. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.
12. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43:491–8.
13. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, *et al.* (2015) SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods*. 12:966–8.
14. Layer RM, Chiang C, Quinlan AR, Hall IM. (2014) LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 15: R84.
15. Cingolani P, Platts A, Wang lL, Coon M, Nguyen T, *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 6:80–92.
16. Tiwari A, Schuiki I, Zhang L, Allister EM, Wheeler MB, *et al.* (2013) SDF2L1 interacts with the ER-associated degradation machinery and retards the degradation of mutant proinsulin in pancreatic beta-cells. *J. Cell Sci.* 126:1962–8.
17. Lorenzon-Ojea AR, Caldeira W, Ribeiro AF, Fisher SJ, Guzzo CR, *et al.* (2014) Stromal cell derived factor-2 (Sdf2): a novel protein expressed in mouse. *Int. J. Biochem. Cell Biol*. 53:262–70.

Cite this article as: Yagil Y, Markus B, Kohen R, Yagil C. (2016) Three interacting genomic loci incorporating two novel mutations underlie the evolution of diet-induced diabetes. *Mol. Med*. 22:560–9.