Mitochondrial and chromosomal genomics in type 2 diabetes

Abhyankar, Avinash

2009

||श्री साई||

Lund University
Department of Clinical Sciences, Malmö
Medical Genetics Unit
University Hospital MAS

Mitochondrial and chromosomal genomics in type 2 diabetes

Avinash Abhyankar, **M.D., M.Sc. (Bioinformatics)**

Academic Dissertation

With the permission of the Medical Faculty of Lund University, to be presented for public examination at Clinical Research Center (CRC) Lecture Hall, Entrance 72, UMAS, Malmö on December 18, 2009 at 1.00 p.m.

Supervisors
Prof. Holger Luthman, Department of Clinical Sciences – Malmö, Lund University, Sweden
Giancarlo Tonolo, MD PhD, Director SC Diabetologia ASL 2 Olbia, Italy

Faculty opponent
Prof. Göran Andersson, Department of Animal Breeding and Genetics (SLU), Uppsala, Sweden

Lund University, Faculty of Medicine

ISSN 1652-8220
ISBN 978-91-86443-14-6
Lund University, Faculty of Medicine Doctoral Dissertation Series 2009:125

Printed by Media-Tryck, Lund University, Sweden
Cover: Diabetes Word Cloud, Generated by Avinash Abhyankar at wordle.net

कर्मण्येवाधिकारस्ते मा फलेषु कदाचन |

मा कर्मफलहेतुर्भुर मा ते संगोस्त्वकर्मणि |

- भगवद्गीता

You have the power to act but, you do not have the power to influence the result.
Therefore, you must act without anticipation of the result, without succumbing to inaction.
- Bhagwad-Gita

Table of Contents

List of publications included in the thesis

1. Hee-Bok Park, **Avinash Abhyankar**, Yuehua Cui, and Holger Luthman
   Interaction between nuclear and mitochondrial genes in the GK rat model for type 2 diabetes. *Submitted*
2. **Abhyankar A**, Park HB, Berger K, Granhall C, Lagerholm S, Åkesson K, Degerman E, Tonolo G, Luthman H
   Novel congenic strain of the Goto-Kakizaki rat reveals nuclear-mitochondrial interactions in type 2 diabetes. *Manuscript*
3. **Abhyankar A**, Hall E, Fathi N, Långberg A, Park HB, Tonolo G, Luthman H
   A gene desert on rat chromosome 9: a candidate for nuclear-mitochondrial interactions in type 2 diabetes. *Manuscript*
4. **Avinash Abhyankar**, Hee-Bok Park, Giancarlo Tonolo, Holger Luthman
   Phylogenetic analysis of mitochondrial DNA in inbred rats. *Submitted*
5. **Avinash Abhyankar**, Hee-Bok Park, Giancarlo Tonolo, Holger Luthman
   Comparative sequence analysis of the non-protein-coding mitochondrial DNA of inbred rat strains. *PLoS One (In press)*

List of publications not included in the thesis

1. M Mezhybovska, Y Yudina, **A Abhyankar** and A Sjölander
   β-Catenin is involved in alterations in mitochondrial activity in non-transformed intestinal epithelial and colon cancer cells. British Journal of Cancer (2009) 101: 1596–1605
2. Tonolo G, Velussi M, Brocco E, Abaterusso C, Carraro A, Morgia G, Satta A, Faedda R, **Abhyankar A**, Luthman H, Nosadini R
   Simvastatin maintains steady patterns of GFR and improves AER and expression of slit diaphragm proteins in type II diabetes. Kidney International (2006) 70: 177-186
3. Nosadini R, Velussi M, Brocco E, Abaterusso C, Carraro A, Piarulli F, Morgia G, Satta A, Faedda R, **Abhyankar A**, Luthman H, Tonolo G
   Increased renal arterial resistance predicts the course of renal function in type 2 diabetes with microalbuminuria. Diabetes (2006) 55: 234-239
4. Nosadini R, Velussi M, Brocco E, Abaterusso C, Piarulli F, Morgia G, Satta A, Faedda R, **Abhyankar A**, Luthman H, Tonolo G
   Altered transcapillary escape of albumin and microalbuminuria reflects two different pathogenetic mechanisms. Diabetes (2005) 54: 228-233

Introduction

This thesis deals with diabetes, more specifically genetic and functional mechanisms of type 2 diabetes in a rat model for type 2 diabetes. Diabetes mellitus is a group of heterogeneous disorders that share the common phenotype of hyperglycemia and glucose intolerance either due to insulin deficiency or impaired insulin action, or both. Currently 285 million people are estimated to have diabetes and the number is expected to be 438 million by the year 2030 [1]. Moreover, an additional 344 million people are estimated to suffer from impaired glucose tolerance with predicted 472 million by the year 2030 [1]. The cost of diabetes to national and individual economies is devastating, largely attributable to complications of the disease [2,3].

Based on the etiology diabetes has been classified into four major types [4].

I. **Type 1 diabetes** (β-cell destruction, usually leading to absolute insulin deficiency)
    a. Immune-mediated
    b. Idiopathic
II. **Type 2 diabetes** (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly secretory defect with insulin resistance)
III. **Other specific types**
    a. Genetic defects of β-cell dysfunction characterized by mutations in:
        i. MODY1: Hepatocyte nuclear transcription factor 4α (HNF4A)
        ii. MODY2: Glucokinase
        iii. MODY3: HNF1A
        iv. MODY4: Insulin promoter factor 1 (IPF1)
        v. MODY5: HNF1B
        vi. MODY6: NeuroD1
        vii. Mitochondrial DNA ←
        viii. Proinsulin or insulin conversion
    b. Genetic defects in insulin action
        i. Type A insulin resistance
        ii. Leprechaunism
        iii. Rabson-Mendenhall syndrome
        iv. Lipodystrophy syndrome
    c. Diseases of the exocrine pancreas (pancreatitis, pancreatectomy, neoplasia, cystic fibrosis, hemochromatosis, fibrocalculous pancreatopathy)
    d. Endocrinopathies (acromegaly, Cushing's syndrome, glucagonoma, pheochromocytoma, hyperthyroidism, somatostatinoma, aldosteronoma)

e. Drug- or chemical-induced (vacor, pentamidine, nicotinic acid, glucocorticoids, thyroid hormone, diazoxide, β-adrenergic agonists, thiazides, dilantin, α-interferon)
f. Infections (congenital rubella, cytomegalovirus)
g. Uncommon forms of immune-mediated diabetes ("stiff-man" syndrome, anti-insulin receptor antibodies)
h. Other genetic syndromes sometimes associated with diabetes (Down's syndrome, Klinefelter's syndrome, Turner's syndrome, Wolfram's syndrome, Friedreich's ataxia, Huntington's chorea, Laurence-Moon-Biedl syndrome, myotonic dystrophy, porphyria, Pader-Willi syndrome)

IV. **Gestational diabetes mellitus (GDM)**

Type 2 diabetes (T2D) is the most common form of diabetes mellitus and results from a complex interaction between genes and environment. The lifetime risk of developing T2D has been estimated of being 38% if one parent had T2D and 60% if both parents had T2D [5,6]. The concordance rates of T2D in monozygotic twins has been shown to be as high as 58% compared to 15-20% in dizygotic twins [7,8], further highlighting the importance of genetic elements in pathogenesis of T2D. The major environmental risk factors include reduced physical activity and overconsumption of calorie-rich diet often leading to obesity and insulin resistance.

To understand the pathophysiology of T2D it is important to understand the physiology of normoglycemia. Insulin, secreted by the pancreatic β-cells, is the principle hormone responsible for regulation of blood glucose and normoglycemia is largely maintained by coordination between insulin action and secretion. Under normal circumstances, the β-cells can compensate changes in insulin action by regulating insulin secretion – if the insulin action is decreased (insulin resistance) the β-cells increase insulin production and vice versa in case of insulin sensitivity. The pancreatic β-cells secrete insulin in response to various secretagogues, glucose being the most prominent. Glucose enters the pancreatic β-cells via the glucose transporters (GLUT1 in rodents and GLUT2 in humans) located in the cell membrane (Figure 1). Inside the cell, glucose is phosphorylated by glucokinase and metabolized to pyruvate which in turn enters the mitochondrion to participate in the TCA cycle ultimately leading to the production of ATP by the respiratory chain. This causes a rise in the cellular ATP/ADP ratio which in turn leads to closure of potassium channels followed by depolarization of the cell membrane and opening of voltage-gated calcium channels. The resulting rise in intracellular calcium leads to exocytosis of insulin located in the storage granules. Mitochondrial ATP production is crucial for glucose stimulated insulin secretion and is abolished by inhibitors of ATP synthesis [9].

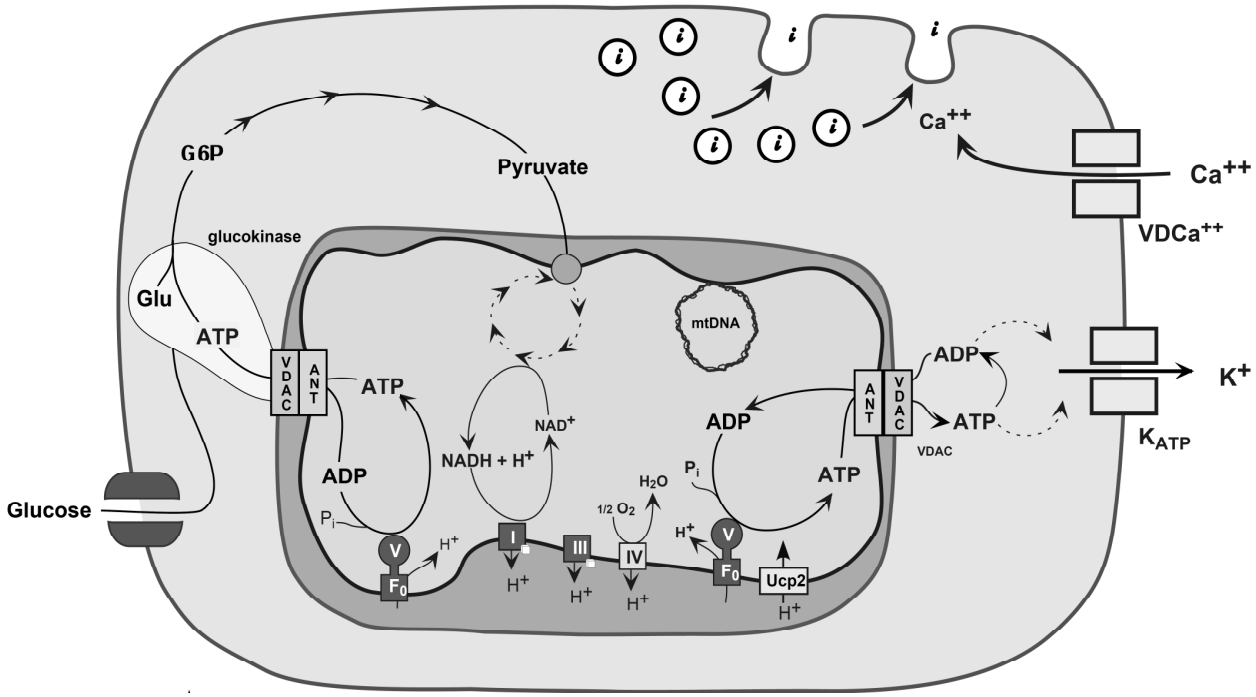

Figure 1. *Glucose stimulated insulin secretion (Image: www.mitomap.org)*

Mitochondria and diabetes

Maternally inherited diabetes and deafness syndrome (MIDD) was one of the first descriptions of mitochondrial diabetes and may account for up to one percent of the total diabetes population [9,10]. Several studies over the last two decades have reported association of diabetes with mutations in the mtDNA [11-20]. Similarly, mitochondrial genes encoded by the nuclear genome have also been implicated in the pathogenesis of diabetes [21-28]. Mitochondria are maternally inherited cytoplasmic organelles, which are essential for generation of cellular adenosine triphosphate (ATP) by oxidative phosphorylation (OXPHOS) in the eukaryotic cell [29]. Mitochondrial ATP is necessary for phosphorylation of glucose by glucokinase and for insulin release by beta cells [30,31]. In addition, mitochondria also host numerous intermediary cellular metabolisms [32] and cell signaling pathways [33]. Therefore, mitochondrial dysfunction is implicated in various types of disease, including metabolic and degenerative disorders [34,35].

Nuclear-mitochondrial interactions

The origin of mitochondria from prokaryotic endosymbionts is strongly supported by comparative genomic sequence data from both prokaryotes and eukaryotes [36,37]. According to the endosymbiotic theory, the separate existence of nuclear and mitochondrial genomes (mtDNA) in eukaryotic cells can be explained by endosymbiont acquisition by the ancestor of eukaryotic cells. Subsequently, intergenomic co-evolutionary processes between the

prokaryotic endosymbionts and their hosts, such as size reduction of the endosymbiont genome (ancestral mtDNA), appear to have occurred in the course of establishment of endosymbiosis [38,39]. Hence, it may not be surprising that mitochondria require coordinated expression of nuclear genome and mtDNA for their proper biogenesis and functions. Most eukaryotic cellular proteins for mitochondrial biogenesis and function are encoded by genes in the nuclear genome although a small number of proteins are encoded by the genes residing in the mtDNA. The vast majority of remaining proteins responsible for generating ATP are encoded by nuclear genome and should be transported to mitochondria [40]. Thus, impairment in the cross-talk between nucleus and mitochondria causing mitochondrial dysfunction are expected to affect the pathogenesis of T2D.

Understanding the genetics of type 2 diabetes

Over the last three decades enormous efforts have been undertaken to understand the genetic basis of T2D with limited success. Until recently, the main approaches to study genetic basis of complex diseases like T2D were based either on genome-wide linkage mapping or association studies of candidate genes. Genome-wide linkage mapping is a hypothesis-free approach to trace patterns of heredity in families with multiple affected subjects in an attempt to locate disease-causing gene variations by identifying traits that are co-inherited. This approach is best used to detect genetic variants with high penetrance or large effect size that rarely contribute to the risk of T2D [41], Table 1. Candidate gene association studies, as the name suggests, try to identify association of chosen 'candidate genes' to a particular trait using case-control or parent-offspring trio samples. The major limitation of this approach is identification of credible 'candidate genes'. Both these approaches need large sample size to make confident detection of variant or gene association. The low-throughput and time consuming genotyping methods made it even more difficult to conduct such studies on a large scale. However, the recent developments in high-throughput genetic screening methods and the availability of large sample sizes have made it possible to conduct affordable powerful studies and identify robust gene-disease associations

Gene nearest to locus	Year	Effect size (odds ratio)
TCF7L2	2006	1.37
KCNQ1	2008	1.29
CDKN2A/2B	2007	1.2
FTO	2007	1.17
HHEX/IDE	2007	1.15
SLC30A8	2007	1.15
THADA	2008	1.15
PPARG	2000	1.14
KCNJ11	2003	1.14
CDKAL1	2007	1.14
IGF2BP2	2007	1.14
NOTCH2	2008	1.13
WFS1	2007	1.12
CDC123/CAMK1D	2008	1.11
HNF1B	2007	1.1
JAZF1	2008	1.1
TSPAN8/LGR5	2008	1.09
ADAMTS9	2008	1.09

Table 1. *Genome-wide significant T2D susceptibility loci.*

[42-54]. The list of identified T2D susceptibility loci is growing at a rapid rate than ever. Table 1 lists some of the genome-wide significant T2D susceptibility loci identified to date. All the susceptibility genes identified to date show only modest effect sizes. Apart from the shortcomings mentioned earlier, studying the genetic basis of complex diseases like T2D in humans is also compromised by the inability to control the environmental factors that heavily influence these diseases. Understanding the effects of mtDNA variants and nuclear-mitochondrial interactions on T2D susceptibility using human population is even more difficult because of the large number of mildly deviant mtDNA haplotypes.

To summarize, even though several genetic and environmental components associated with T2D have been identified there is still a lack of comprehension in the understanding of the molecular mechanisms involved. Studying a complex disease like T2D, that is a result of interactions between the genetic (nuclear and mitochondrial) and environmental components, is extremely difficult especially in outbred populations like humans. Inbred animal models on the other hand offer a more standardized alternative to study such interactions.

The laboratory rat

Rattus norvegicus is a commonly used experimental animal that has been extensively used to study several important disease traits [55-62]. As early as 1909 the first inbred rat strain, named PA, was derived from albino rats at the Wistar Institute in Philadelphia [63]. However, non-inbred colonies of black hooded rats have been used since 1856. Hundreds of genetically defined inbred strains representing a range of human disease traits are currently available for research. Completion of the rat genome sequence was a major step forward in the quest for understanding the molecular basis of disease traits in these models [64]. The Rat Phenome Project has made available a wide range of phenotypic data for over 200 different rat strains [65]. Moreover, genomic variation (SNP) data for over 300 different rat strains is now available [66]. Until recently only 14 complete rat mtDNA sequences were available in public databases. We have added 13 additional complete mtDNA sequences to this repertoire. For the laboratory rat we now have the complete sequence of the nuclear genome, phenotypic measurements, variation data and a fairly large number of mitochondrial genome sequences at our disposal. This excellent combination of data is a powerful tool in understanding disease mechanisms and designing models to test new hypotheses.

Aims

1. To investigate whether interaction between the genetic variants in nuclear and mitochondrial genes affect the susceptibility to type 2 diabetes in an animal population segregating two different forms of mtDNA
2. To characterize the *Niddm71* locus using a congenic strain (C9B) derived from the normoglycemic F344 and the diabetic GK rat
3. To identify the genetic elements encoding the defects in insulin secretion and mitochondrial function in the congenic C9B rat
4. To analyze the mitochondrial genomes of various inbred rat strains in an attempt to identify genes or variable sites that might affect the mitochondrial function

Approach

Study 1: As the first step, we wanted to investigate the effects of nuclear-mitochondrial interactions on the susceptibility of type 2 diabetes. We arranged a reciprocal intercross (a pair of crosses) that resulted in an F2 population with divergent mtDNA originating either from the diabetic GK rats (Cross 1) or the normoglycemic F344 rats (Cross 2). Mitochondrial genomes of these two strains differ at 110 nucleotide positions. Figure 2 shows a diagrammatic representation of the breeding scheme employed for generation of the two crosses.

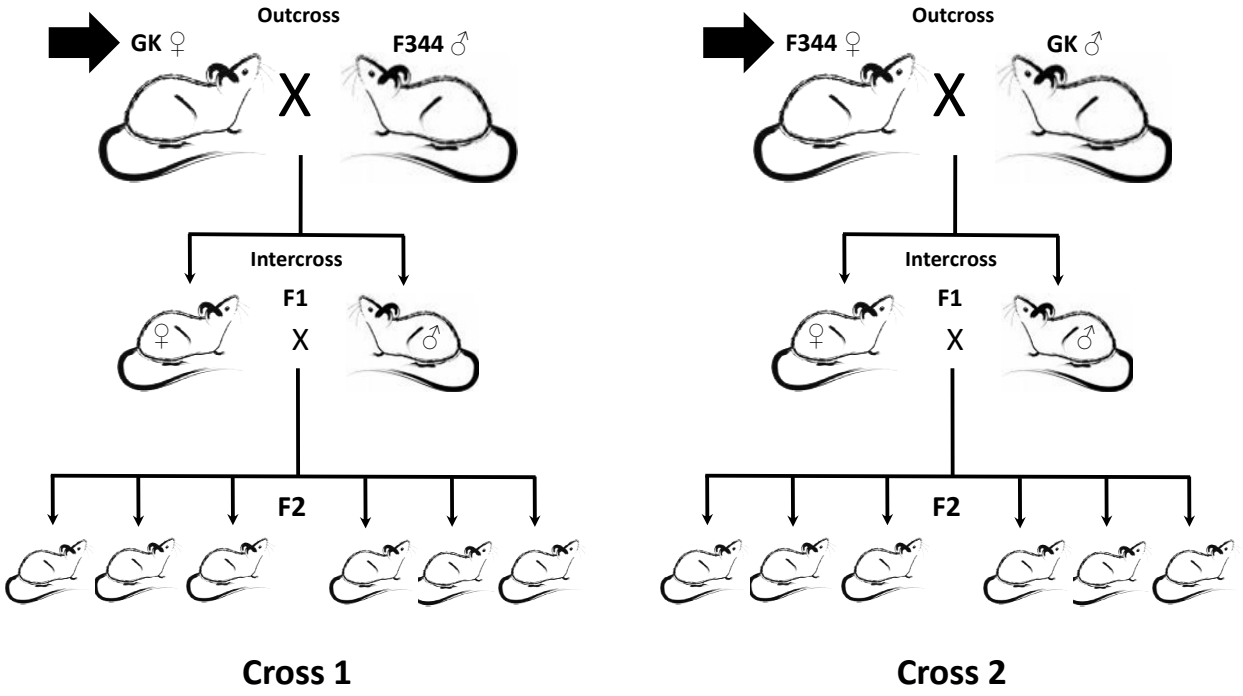

Figure 2. *Breeding scheme for the F2 reciprocal intercross populations*

As can be seen from the figure, the only difference between the two crosses is the founder maternal strain – diabetic GK in Cross 1 and normoglycemic F344 in Cross 2 (indicated by black arrows). Thus, both the crosses share an identical nuclear DNA sequence while Cross 1 harbors mtDNA of GK origin, Cross 2 harbors mtDNA of F344 origin. This experimental setup allowed us to investigate a scenario that is impractical to test in human populations. Genome-wide quantitative trait locus (QTL) analyses led to identification of nuclear loci that affected glucose or insulin traits depending on the mtDNA genotype. The statistical significance of interaction between QTL and mtDNA was evaluated by likelihood ratio test. Loci on chromosomes 2 and 9 showed genome-wide significant linkage to fasting glucose and fasting insulin respectively. Additionally, an epistatic interaction between loci on chromosomes 4, 14 and mtDNA was also linked to postprandial glucose.

Study 2: Of the loci identified, we chose the locus on chromosome 9 (*Niddm71*) for further investigation, since it displayed the largest difference in the LOD scores for fasting insulin between Cross 1 and Cross 2. To dissect this locus further we established a congenic strain, C9B, which harbors approximately 7 Mb of homozygous GK genome corresponding to the *Niddm71* locus on a homozygous F344 background and F344 mtDNA. Figure 3 shows a diagrammatic representation of the breeding scheme employed for generation of the congenic strain. To ensure transmission of F344 mtDNA in the resulting congenic rat, the breeding was initiated with a cross between normoglycaemic F344 female and diabetic GK male. This was followed by backcrossing the resulting male progeny to the parental F344 female. Only those rats harboring desired GK alleles were selected for each round of backcrossing. After twelve rounds of backcrossing, the generation N12 was named C9B and has been maintained by brother-sister mating. This C9B strain is homozygous for GK from 35.5 Mb to 42.2 Mb on chromosome 9 on a homozygous F344 genetic background with mtDNA and chromosome X from F344 and the Y chromosome from GK. To avoid effects of the estrous cycle, only male rats were used for functional characterization. Glucose homeostasis was assessed by performing intraperitoneal glucose tolerance test (IPGTT) and in-vitro insulin release in pancreatic islets. Mitochondrial function was assessed by monitoring the oxygen consumption with a Clark oxygen sensor electrode. Expression levels of RNAs for key genes involved in mitochondrial function were measured. With the help of these assays we were able to show that the *Niddm71* locus codes for a defect in insulin secretion and a defect in mitochondrial function.

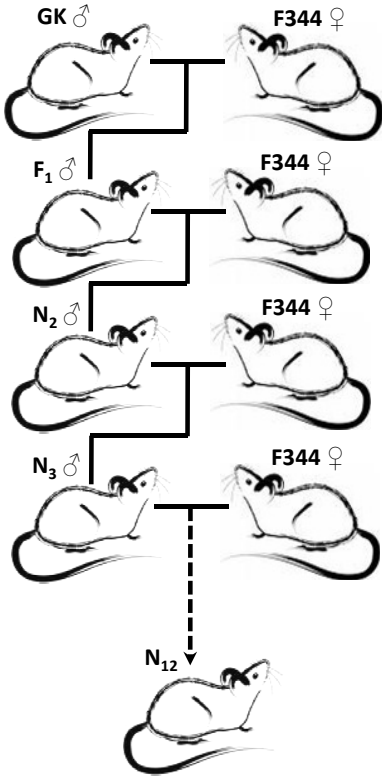

Figure 3. *Breeding scheme for generation of the congenic rat - C9B*

Study 3: Having identified and characterized the *Niddm71* locus on chromosome 9, the next step was to identify the genetic elements underlying the encoded phenotype. Utilizing the information derived from single marker QTL analysis we performed multiple marker QTL analysis to account for multiple linked locus effects[67,68] and were able to narrow down the region of interest from 8.9 Mb to 4.7 Mb. With the rapid progress in various molecular biology techniques and open access policies we have access to enormous primary and inferred biological data. We used information available at Ensembl database, PubMed, OMIM[69], OMIA[70], GeneCards[71], iHOP[72], STRING[73], CDD[74], GEO[75] to extract the known information and predict functions of unknown genes. In the third study we demonstrate how information available at

public biological databases could be used to prioritize genomic elements and reduce unnecessary laboratory work, saving valuable time and money.

Studies 4 & 5: The first three studies deal with the nuclear component of nuclear-mitochondrial interactions. In the last two studies we describe analyses of genetic variation in the mitochondrial genome. Sequence analysis of the complete mtDNA of F344 and GK rats revealed 110 variable nucleotide positions between them making it difficult to identify the position/s that might affect mitochondrial function. To simplify the task we sequenced additional eleven complete mtDNA from various inbred rat strains. These, along with 14 publicly available rat mtDNA sequences were used to perform comparative and evolutionary analyses. Based on the results we were able to identify several mtDNA encoded genes and sites within the non-coding D-loop that might play an important role in mitochondrial function.
To assess the collective effect of F344 and GK mtDNA haplotypes we also established conplastic strains. This experimental setup allowed assessing how mitochondrial function and glucose homeostasis is affected by different combinations of nuclear DNA and mtDNA.

To summarize, we

1. used QTL analyses to identify nuclear loci involved in nuclear-mitochondrial interactions
2. established a congenic strain representing one of the identified locus
3. used various bioinformatics tools to prioritize genetic elements within the locus
4. performed comparative and evolutionary analyses on the mtDNA to identify the mitochondrial genes that might play an important role in mitochondrial function
5. assessed the effect of F344 and GK mtDNA haplotypes on glucose homeostasis and mitochondrial function using conplastic strains

Results

Analysis of the nuclear genome

Study 1

Genome-wide QTL analyses for glucose and insulin traits were performed with adjustment for maternal environmental effects. Single QTL analysis was performed on each cross separately and also using the combined Cross1+Cross2 dataset. If a QTL was detected in one cross, the LOD score at the corresponding position in the other cross was considered and the difference between these two scores (ΔLOD score) was calculated to identify cross-specific QTLs. Based on the ΔLOD scores, we identified four cross-specific genome-wide significant QTLs namely - *Niddm2* (chromosome 2, fasting glucose), *Niddm1* (chromosome 1, glucose at 60 min), *Niddm71* (chromosome 9, fasting insulin) and *Niddm71* (chromosome 9, insulin at 30 min). Likelihood ratio analysis was performed for these four QTLs to assess interaction between QTL and cross. Accordingly only three of them provided such evidence (Figure 4). The *Niddm71* QTL for fasting insulin also had the highest ΔLOD score (Figure 4B).

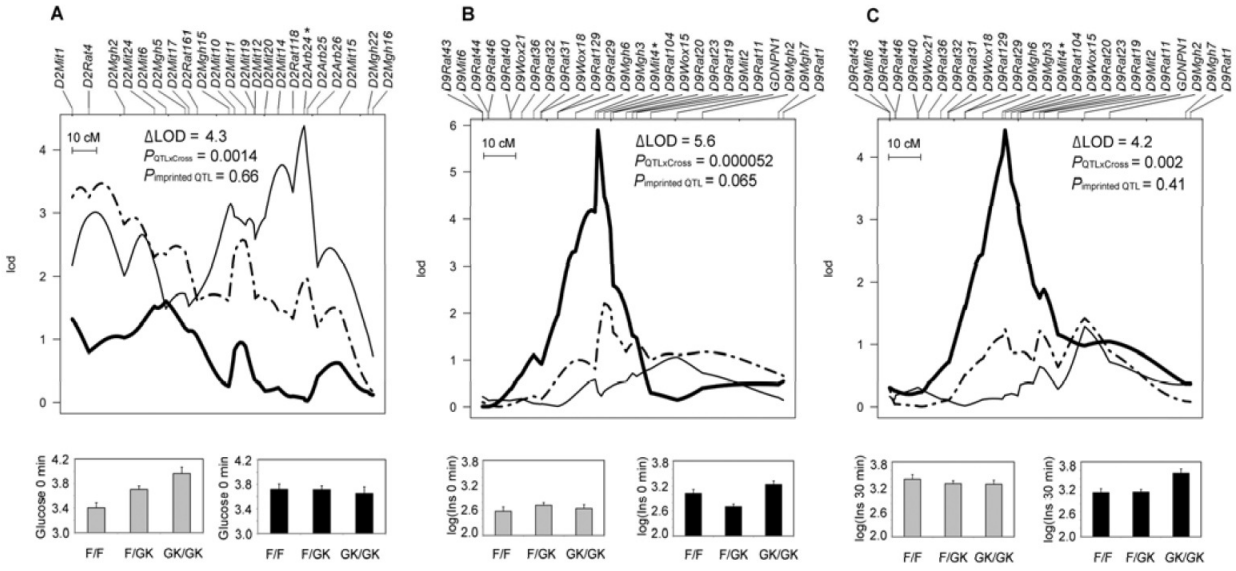

Figure 4. *Genome-wide significant cross specific QTLs showing significant QTL x Cross interaction. Thin lines represent QTL profiles for Cross 1, thick lines represent QTL profiles for Cross 2 and broken lines represent QTL profile for the combined dataset. A: Niddm2 (fasting glucose), B: Niddm71 (fasting insulin), C: Niddm71 (insulin at 30 min).*

We also identified a cross-specific epistatic QTL pair for glucose at 120 min. In Cross 1, loci *D4Mit28* (chromosome 4) and *D14Mit10* (chromosome 14) did not reveal significant effects independently but, together, they exerted significant effect on the glucose trait through interaction with cross. Within all these cross-specific QTLs we identified 40 genes that are important to mitochondrial function with high probability.

Study 2

Niddm71, the cross-specific QTL for fasting insulin with the highest ΔLOD score, was chosen for further investigation. A congenic strain, C9B, harboring the diabetic GK alleles for *Niddm71* and Y chromosome on a normoglycemic F344 nuclear and mitochondrial genetic background was established. Comparison of phenotypes for various traits was performed between the congenic C9B and normoglycemic F344 strains. During the course of IPGTT, plasma glucose and insulin were measured after 6 hours of fasting followed by measurements at 20, 40 and 60 minutes after intra-peritoneal glucose injection. At all these time points, C9B had significantly lower insulin levels as compared to F344 pointing towards a defect in glucose stimulated insulin secretion. To confirm this finding we assessed the glucose stimulated insulin response in isolated pancreatic islets. Insulin secretion was measured after incubating the islets at low (3.3 mM) and high (20 mM) glucose concentrations. At low glucose concentration there was no difference in insulin secretion. However, at high glucose concentration the C9B islets secreted approximately 40 percent less insulin as compared to F344 confirming the defect in glucose stimulated insulin secretion. To assess the mitochondrial function we analyzed the cellular oxygen consumption and measured the RNA expression levels of key genes involved in mitochondrial biogenesis and function. Oxygen consumption measured at the ages of six and twelve weeks showed clearly significant difference of mitochondrial function between C9B and F344 strains. The C9B strain had marked reduction in oxygen consumption – approximately 50 percent lower at six weeks and 40 percent lower at twelve weeks of age. It is noteworthy that there were no overlapping measurements between the C9B and F344 strains. We also measured RNA transcript levels of four key mitochondrial genes – Ppargc1a, 16S rRNA, ND2 and ND6. We found that the Ppargc1a RNA levels are up-regulated in C9B but the three mitochondrial genes (16S rRNA, ND2 and ND6) are down-regulated. Ppargc1a is a major regulator of mitochondrial biogenesis and has been shown to be affected in T2D patients [25]. The remaining three genes are encoded by the mtDNA and are expected to respond to changes in Ppargc1a transcript levels. Our results indicate a break in the communication downstream Ppargc1a in the C9B strain.

Study 3

Characterization of the congenic C9B strain revealed that *Niddm71* harbors genomic regions that influence glucose stimulated insulin secretion and mitochondrial function. To further dissect this locus we used a comprehensive bioinformatics and experimental approach. Based on the results of single-marker QTL analysis (from Study 1) we performed multiple-marker QTL analysis to account for linked locus effects and were able to narrow down our search to a 4.7 Mb region flanked by markers *D9Mgh3* and *D9Rat104*. This region has 36 genes and an approximately 2 Mb region that does not code for any protein gene (gene desert). We

started with analyzing the SNP information available for different rat strains. After collecting the SNP information for 637 SNPs from 388 strains we generated haplotype sequences of these SNPs for each strain. Since most of the strains were crosses between different inbred strains we used data from selected 115 strains. A phylogenetic tree was generated using the haplotype sequences for *Niddm71*. Surprisingly, these 115 strains grouped only in two major clades – a large clade comprising of 105 strains and a small clade comprising of only 10 strains. All the strains in the small clade display some form of phenotype with metabolic origin. Consider a typical identical-by-descent (IBD) scenario. Assume that, to start with, the alleles responsible for the phenotype displayed by the congenic C9B arise by mutations at a particular genetic locus. Several generations and crossovers later, this locus segregates but the alleles responsible for glucose homeostasis trait are still in linkage disequilibrium. This group of animals might be one explanation to why only 10 strains sharing a common phenotype make a small group and the remaining 105 strains a large group. We then sought out which of the SNPs used for the phylogenetic analysis are responsible for this specific grouping and identified three SNPs that could explain this pattern. Using positional information about these SNPs we identified three blocks (Figure 5) and focused our attention on the genetic elements within these blocks. The identified blocks have 18 genes and approximately 1Mb of gene desert. For the 18 genes we extracted all known information from public biological databases. Based on this information, we further selected only nine of these for functional analysis. Analysis of RNA levels for these nine genes in pancreatic islets revealed that five of the nine genes have more than two-fold expression difference between the normoglycaemic F344 and the diabetic GK rats. Identifying several differentially regulated genes was unexpected and pointed towards a common regulatory mechanism for these genes. The gene desert had the possibility of harboring such regulatory regions. Evolutionary conservation analysis of the gene desert led to identification of several genomic segments that were highly conserved between species ranging from human to fish. This high degree of conservation signifies the evolutionary and functional importance of these regions. Sequence analysis of these conserved loci between the F344 and C9B strains

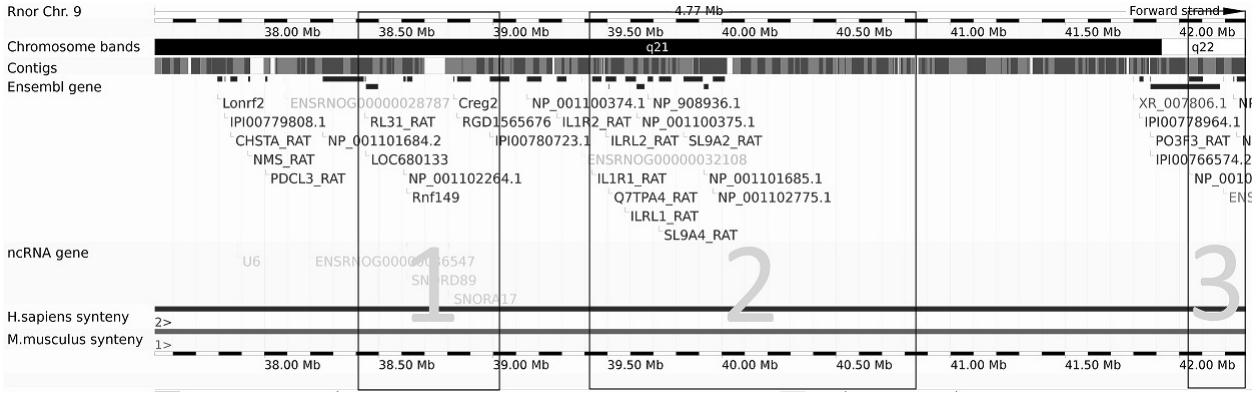

Figure 5. *Location of haplotype blocks identified using SNP analysis (Image: www.ensembl.org)*

revealed several variable nucleotide positions that led to either gain or loss of binding sites for transcription factors implicated in diabetes and mitochondrial function.

To summarize, we identified several genes that are differentially expressed between the normoglycaemic F344 and diabetic GK rats. We have also identified few highly conserved non-protein-coding regions that might possibly play a role in the regulation of transcript levels for these genes.

Analysis of the mitochondrial genome

We sequenced the complete mitochondrial genomes of F344 and GK rats and found 110 nucleotide differences between them. Characterization of conplastic strains (F344-mtGK and GK-mtF344) revealed that changing the combination of nuclear and mitochondrial genomic haplotypes influenced not only mitochondrial function but also glucose homeostasis (preliminary data, not shown). The large amount of variation between F344 and GK mtDNA made it difficult to identify the functionally important variants. Hence, we sequenced complete mtDNA from eleven more commonly used inbred rat strains. For the comparative analysis we also included 14 complete rat mtDNA sequences available in public databases. Since the protein-coding and the non-protein-coding regions of mtDNA serve different purposes – function and regulation of function – the variation pattern and the evolutionary pressures are expected to be different. Furthermore, the relative significance of coding sequence variation compared to the regulatory sequence variation, from an evolutionary perspective, remains poorly understood [76]. For this reason we investigated the protein-coding and the non-protein-coding regions separately. Study 4 describes analyses of the protein-coding region while Study 5 describes the analyses of non-protein-coding region of mtDNA.

Study 4

Comparison of the protein-coding region of 27 mtDNA sequences revealed 194 variable sites. Of these 90 were unique either to the reference BN or any of four wild rat sequences. Excluding the singletons, there were 104 variable nucleotide sites, 31 of which led to an amino acid change (non-synonymous change). Highest numbers of such amino acid changes were observed in the ND2 and ND4 genes. Neighbor-joining phylogenetic tree confirmed the known relationships between the strains. Two major clades could be identified - 'Wistar clade' consisting of strains originating from the rat strain coming from the Wistar Institute in Philadelphia and a second clade consisting of strains not originating from the Wistar rat. To test whether the protein-coding region is evolving randomly (neutrally) or by a non-random process we used three different tests – the Tajima's D test, Fu & Li's D and F tests, and likelihood ratio tests (LRT). Four genes (ND2, ND4, COII and Cytb) showed statistically significant deviations from neutrality. Using LRTs we identified individual sites (codons) that are experiencing selection. Non-neutral sites were identified in three genes encoding subunits of complex I (ND2,

ND4 and ND6) and one gene each from complex III (Cytb), complex IV (COII) and complex V (ATP6). Six of the identified sites were variable within the Wistar group of strains - A18V, S150N, T265A in the ND2 and T356A, I401V, L419P in the ND4 were variable within this group. These six sites were further evaluated using comparative analysis. Conservation index (CI) was calculated based on the conservation status in 41 different mammalian species. Possible functional effects of these sites were also predicted by calculating the subPSEC scores using the web based PANTHER classification system [77,78]. Amino acids N150 in the ND2 gene and A356 in the ND4 gene had the highest CIs of 85% and 65% respectively. The subPSEC scores for the six sites variable within the Wistar-group were ND2 - A18V:- 2.102, S150N: -3.068, T265A: -3.997, ND4 - T356A: -4.812, I401V: -3.224, L419P: -5.677.

In conclusion, we identified several genes and sites within them that are experiencing selection and, thus, might be important for mitochondrial function.

Study 5

Proper function of mammalian mitochondria necessitates a coordinated expression of both nuclear and mitochondrial genes. Non-protein-coding regions of the mtDNA form a major component of this regulated expression. The mitochondrial ribosome (mitoribosome) is composed of a small subunit consisting of 12S rRNA and 29 proteins and a large subunit consisting of 16S rRNA and 58 proteins [79]. Within the 27 rat mtDNA sequences seven variable positions were observed in the 12S rRNA and 24 variable positions in the 16S rRNA. None of the variable positions within the 12S rRNA alter the predicted secondary structure. However, the variant positions in the 16S rRNA altered the predicted secondary structure to a major extent. Only one position in the 16S rRNA (position 2170) was conserved among nine mammalian species included in the study and was located in close proximity to the L1-binding domain of 16S rRNA. The L1-binding domain is highly conserved and has a dual function - ribosomal protein binding and translational repression. Of the 22 tRNA genes coded by the mtDNA only five had variable sites. Three variable sites were observed in *tRNA-Cys*, two in *tRNA-Pro*, while *tRNA-Tyr*, *tRNA-Asp* and *tRNA-Thr* had one variable site each. The D-loop is the only major non-coding region of mtDNA and harbors sequences important to regulation of mtDNA transcription and replication. Thirteen variable sites were found in the D-loop – eleven substitutions and two insertion/deletions. We mapped these variations to the known D-loop functional sites [80-83]. Seven substitutions and one insertion/deletion were located in these functional sites. We also tested the non-protein-coding regions for non-neutral evolution but did not find any support for such processes.

In conclusion, we have identified sites in the non-protein-coding region of rat mtDNA that likely play a role in regulation of mtDNA transcription and replication.

Discussion

One of the first descriptions of T2D dates back to 600 BC when the Indian physician Sushruta described a disease, he called, madhumeha (sweet urine) and linked the disease to 'injudicious diet', sedentary lifestyle and obesity. The recommended treatment included dietary changes and exercise to minimize its consequences [84]. But, after so many decades of intense research with powerful tools at our hand we still lack a complete understanding and a cure for this disease. This statement does not highlight lack of progress but underscores the complexity of T2D. Considering this complexity, the success that has been achieved till date is in fact remarkable. Although environmental factors (e.g. overeating, physical inactivity) are known to trigger T2D, genetic factors (nuclear and mitochondrial) are equally necessary components of the disease. These factors make it difficult to design and undertake a human study that will efficiently and accurately dissect the pathophysiological basis of T2D. However, with the advent of newer high-throughput genotyping technologies it has become possible to undertake large-scale human studies that have led to identification of several genetic variants that show a modest effect on the susceptibility to T2D [42-54]. Several genes within the vicinity of these identified loci have shed novel insights into disease mechanisms. Moreover, nuclear and mtDNA mutations leading to mitochondrial dysfunction have been implicated in the pathogenesis of this disease. Considering the complex interactions between the nuclear genome, mitochondrial genome and environmental factors, translating the insights to understanding the molecular basis of T2D will not be possible without the use of animal models.

To provide evidence of nuclear-mitochondrial interaction, multiple lines of approaches have been proposed. Artificial human cell lines were constructed with the same nuclear genome but different mtDNA to evaluate effects of the interaction on respiratory capacities at cellular the level [85]. Systematic and integrative screenings of nuclear-encoded mitochondrial proteins and genes were performed using various experimental systems [86-89], which provided a comprehensive list of mitochondrial proteins encoded by the nuclear genome. These lists can be used to identify nuclear-encoded candidate genes interacting with mitochondria and in combination with previously mapped genetic intervals implicated in mitochondria associated disorders [90,91]. Furthermore, these lists can directly be combined with homozygosity mapping to identify positional candidate genes causing rare forms of mitochondrial disease [92,93]. Knockout mice were used to elucidate the roles of nuclear genes encoding mitochondrial proteins in OXPHOS disorders [94,95]. This approach also demonstrated nuclear regulators of mitochondrial transcription initiation [96,97]. However, these knock-out mice experiments do not represent common multifactorial disorders associated with mitochondrial dysfunction caused by naturally occurring genetic variations. To investigate effects of interaction between the nuclear genome and mitochondria on phenotypes of interests, conplastic strains can be generated by continuous backcrossing resulting in a new strain with mtDNA from another strain. This

approach was applied for the genetic dissection of cognition in mice [98]. Conplastic strains in rats and mice were also used to provide evidence for the effect of mitochondrial genetic variations on complex quantitative phenotypes [61,99]. Alternatively, Rand et al. reported effect of interaction between nuclear genetic background and mtDNA on longevity using reciprocal crossing of two fruit fly strains carrying divergent mtDNAs [100]. Using reciprocal backcross in mice, Johnson et al. showed an interaction between nuclear quantitative trait locus (QTL) and mtDNA mutation in hearing loss [101]. However, we still have very limited information about the interactions of individual nuclear loci with mtDNA investigated in the frame of QTL analysis. In addition to individual effects of mitochondrial-encoded genes and nuclear-encoded genes interacting with mitochondria, other factors, such as sex chromosomes, genomic imprinting, or maternal environments, are likely to be involved in the inheritance of mitochondria as well as phenotypic expression of nuclear-mitochondrial interactions. Together these multiple factors make it complicated to identify QTLs × mtDNA interaction and to investigate contribution of the interaction to the phenotype.

Animal models provide us with complex living systems that can be maintained under very controlled conditions – genetic and environmental - making them ideal for studying molecular mechanisms of complex diseases like T2D. Several rodent models of T2D have been developed and successfully used to dissect the genetic basis this disease [55-60,62,102-107]. The GK rat is one of the most studied rat models of spontaneous T2D [57]. Generated by selective breeding of non-diabetic Wistar rats that were glucose intolerant after oral glucose challenge, these rats become diabetic very early in life. To understand the role of nuclear-mitochondrial interactions in T2D we setup a reciprocal cross between the diabetic GK and normoglycaemic F344 rats. This reciprocal cross consisted of two individual crosses that shared identical nuclear DNA haplotypes while they differed in the mtDNA haplotypes at 110 nucleotide positions. We performed cross-separated linkage analysis to map QTLs segregating only in the presence of specific mtDNA genotype, which is a first-line indication for the interactions between nuclear QTL and mtDNA. Two genomic loci were found linked to three traits depending on the mtDNA haplotype. Within these loci we were able to identify 40 nuclear-encoded mitochondrial genes. Reciprocal-cross setup has earlier been used basically to understand the effect of maternal factors on various traits [108-111]. The maternal factors include effect of the sex chromosomes, the mtDNA, genomic imprinting, the intrauterine environment experienced by the fetus during its development. It is seemingly straightforward to sequence the 16300 base pairs long mtDNA and identify the nucleotide differences between the parental strains. However, assessing the genome-wide imprinting pattern is not as standardized and technologically simplified as cycle sequencing. Several approaches have been used to make a genome-wide assessment of the imprinting patterns [112-117]. In the present study we used a statistical approach, imprinted QTL analysis, to test if genomic imprinting can explain the observed cross specificity at the identified QTLs but did not find any support for the same. However, based on statistical evidence we

cannot completely rule out imprinting and experimental verification of imprinting at the identified loci will allow more specific conclusions in the future.

Transforming the identified loci into mechanistic understanding of a disease is an arduous task. Even though we identified several loci conferring susceptibility to T2D, the resolution provided by F2 intercross is not enough for positional cloning of a particular gene or genomic element. Congenic animals could be used to narrow-down and characterize the precise phenotype encoded solely by the identified locus [118-123]. We decided to further dissect the *Niddm71* locus that displayed the maximum ΔLOD score between the two crosses and generated a congenic rat that harbors diabetic GK alleles at this locus on a normoglycaemic F344 genetic background. This congenic rat, C9B, was generated by twelve rounds of backcrossing followed by inbreeding. Theoretically, this means that, apart from the Niddm71 locus, the C9B rat still retains approximately half million base pairs of GK genome. The presence of such unknown GK alleles randomly spread throughout the nuclear DNA was considered while interpreting the characteristics of C9B rat. The insulin trait observed in F2 intercross for the *Niddm71* locus was maintained in the congenic rat. To investigate whether this locus harbors genetic elements that affect mitochondrial function, we measured the cellular oxygen consumption and RNA expression levels of selected genes. Three genes encoded by the mtDNA (16S rRNA, ND2, ND6) and one nuclear-encoded gene (Ppargc1a) were chosen. Ppargc1a has been shown to be a major regulator of mitochondrial biogenesis [25]. The genes for 16S rRNA and ND2 are encoded by the mtDNA heavy strand while ND6 is encoded by the light strand. The mtDNA is transcribed as a polycistronic transcript, one each for the heavy and light strands. However, for the heavy strand, a third truncated transcript encompassing the two ribosomal genes is also generated at rates different than that for the complete heavy strand transcript [124]. The three chosen mitochondrial genes represent all these three transcripts while Ppargc1a would measure the need for mitochondrial biogenesis. Our results suggest that the C9B rat has defective mitochondrial function measured in the form of reduced oxygen consumption and a break in the nuclear-mitochondrial cross-talk evident from the lack of mtDNA transcriptional response to Ppargc1a. Since Ppargc1a RNA levels were increased in C9B, the lower expression of mitochondrial genes point towards a defect downstream Ppargc1a that might lead to reduced oxygen consumption and insulin response seen in the C9B rat. It is likely that this downstream factor is encoded by *Niddm71* locus that has 36 genes and a fairly large gene desert. We speculate that this downstream factor might either be one or more of the 36 genes or a sequence in the gene desert that might regulate transcription of one or several of these genes. Studies conducted on Ppargc1a knockout mice showed that muscle specific heterozygotes had a delay in insulin release during a glucose tolerance test [125]. This cross-talk between the muscle and pancreatic β-cell might involve IL-6 that is secreted by different metabolic tissues (including skeletal muscle) and exert effect on insulin secretion and sensitivity [126,127]. It has also been shown that there is extensive cross-talk between different cytokines,

especially IL-1 and IL-6, in T2D [128]. The *Niddm71* locus harbors a interleukin receptor gene cluster with six IL-1 related genes. Two of these genes, IL1R1 and IL1R2, serve as IL-1 receptors and exert completely opposite effects on IL-1 signaling. Larsen *et. al.* showed that blockade of IL-1 using anakrina (IL-1 receptor antagonist) led to improvement of glycemia and beta-cell secretory function [128]. Binding of IL-1 to IL1R1 leads to relay of signal while binding of IL-1 to IL1R2 blocks this IL-1 signaling. Overexpression of IL1R2 will truncate the IL-1 signaling, probably similar to anakrina leading to beneficial effects on insulin secretion, while overexpression of IL1R1 might lead to opposite effects as displayed by the C9B rat. Comprehensive functional analysis of these genes in various insulin target tissues of C9B will elucidate the exact role of this gene cluster in insulin secretion and mitochondrial function. We performed RNA expression analysis for ten genes within the *Niddm*71 locus and found that more than half of these genes are at least two-fold overexpressed in the pancreatic islets of diabetic GK rat. This finding led us to investigate the role of the gene desert in transcriptional regulation of these genes. Even though no protein-coding genes were identified within the *Niddm*71 gene desert, it might still harbor functionally important features. Conservation profile deduced from comparative sequence analysis of distantly related species gives a clue about evolutionary, and hence functional, importance of a genome region. If a locus is functionally important, it will be conserved across different species. We identified several regions within the *Niddm71* gene desert that were highly conserved all the way from humans to fish. Sequence analysis of these regions in GK and F344 DNA revealed several variable nucleotide positions that modify predicted binding sites for several key transcription factors implicated in diabetes and mitochondrial function. Few other studies have reported identification disease modifying loci within gene deserts and attempts have also been made to elucidate the functional features within the gene deserts [129-135]. The recent genome-wide association studies for T2D have identified several loci that lie within gene deserts. Detailed studies need to be undertaken in order to understand their contribution to T2D susceptibility.

The mtDNA encodes few but essential components of the respiratory chain that produces most of the cellular ATP crucial for insulin secretion and sensitivity. Our study aimed at identification of cross-specific QTLs was based on a reciprocal cross between diabetic GK and normoglycaemic F344 rats. The reciprocal cross was set up to result in two crosses with identical nuclear genome sequence but different mitochondrial genomes. The cross-specific QTLs thus identified were a result of interaction between the nuclear locus and the cross effect (probably represented by the different mtDNA haplotypes). The mtDNA of F344 and GK strains differ at 110 nucleotide positions. This large amount of variation makes it difficult to identify the pathogenic mutation/s. The comparative analysis principle we used to identify conserved regions within the gene desert could also be applied to the mtDNA. However, for the mtDNA, the aim was to identify functionally important nucleotide variants. For this reason, we made comparative sequence analysis between different *Rattus norvegicus* inbred strains and four

sequences from rats caught in the wild. To identify functionally important regions and sites we used different tests for selection. Evolutionary processes acting on the protein-coding genes are targeted towards modifying the protein function, while those acting on non-protein-coding genes and regulatory regions are targeted towards regulation of expression of gene products. Therefore, the underlying evolutionary mechanisms are expected to be different for the two types of sequences. Moreover, the non-coding D-loop and the coding regions of the mtDNA have different substitution rates [136,137]. Hence, the main aim was to investigate the presence and mechanisms of selective forces in the non-protein-coding genes and the D-loop that would reflect their functional significance. Selection has been operating ever since the first rat strains were domesticated, e.g., as a result of being kept in captivity with its different flora of microbial pathogens, special requirement for high fecundity, large sibships, low aggressivity and calm easy behavior in general. The adaptation to different environmental conditions in different animal houses is unlikely to be convergent. Moreover, several of the strains belonging to the Wistar-group are models for metabolic diseases. These strains have been subject to strong phenotype selection prior to inbreeding. Since these strains have been selected for an extreme phenotype and have survived with high fitness for several generations with this un-physiological phenotype, we expected some form of selection. Six amino acid changing mutations were observed within the Wistar-derived strains. These were two type 2 diabetic GK strains, three hypertensive strains MHS/Gib, SHR/Mol, GH/OmrMcwi and the hypertensive stroke-prone SHRSP. Even though the F344 containing clade has strains derived from different founders no variation was observed within this group for all the six positions. These positions were also fully conserved in the wild population. It is therefore quite possible that these mutations are a result of adaptation to an extreme metabolic phenotype. Emergence of these 'young' mutations only in the strains that have been subject to a phenotype selection process before inbreeding suggests that major phenotypic perturbations leads to adaptive changes at the mitochondrial genomic level. Since mtDNA has a mutation rate almost ten times higher than the nuclear genome, it will be the site where the first adaptive changes will be possible and evident. By analyzing the inbred rat strains we are looking into only a century of evolution. What we present here might be the first signs of such adaptive evolutionary changes in response to survival with extreme phenotypes.

Summary and Conclusions

To understand the role of nuclear-mitochondrial interactions in T2D we arranged a reciprocal cross between the diabetic GK and the normoglycaemic F344 rat strains. This reciprocal cross consisted of two individual crosses that shared identical nuclear DNA sequences while they differed in the mtDNA haplotypes at 110 nucleotide positions. We performed cross-separated linkage analysis to map QTLs segregating only in the presence of specific mtDNA genotype, which is a first-line indication for the interactions between nuclear QTL and mtDNA. Two genomic loci were found linked to three traits depending on the mtDNA haplotype. Within these loci we were able to identify 40 nuclear-encoded mitochondrial genes.

The *Niddm71* locus displayed the highest LOD score difference between the two crosses and was dissected further using a congenic rat, C9B, which harbors diabetic GK alleles at this locus on a normoglycaemic F344 genetic background. Intraperitoneal glucose tolerance test (IPGTT) and tests on isolated islets showed that C9B had impaired glucose stimulated insulin response. To assess the mitochondrial function we investigated oxygen consumption in adipose tissue using an oxygen sensitive Clark electrode. C9B had significantly reduced oxygen consumption as compared with F344. We further determined the RNA expression levels for three mitochondrial genes (16S rRNA, ND2, and ND6) and Ppargc1a in various insulin target tissues. We found that, in C9B as compared to F344, Ppargc1a RNA levels were significantly higher while the levels for all mtDNA coded genes were significantly lower implying a break in nuclear-mitochondrial communication downstream Ppargc1a. In summary we show that C9B has impaired glucose tolerance, insulin secretion, mitochondrial function and nuclear-mitochondrial communication. With a combination of bioinformatics and experimental methods we identified several genes and non-coding sequences within the *Niddm71* locus that likely play a role in insulin secretion and mitochondrial function.

In the first three studies we have shown evidence for mitochondrial genotype-specific modification of nuclear QTLs and a break in nuclear-mitochondrial communication resulting from interactions between one identified nuclear locus of GK origin and F344 mitochondrial genotype. To investigate the genetic variation between the F344 and GK mitochondrial genomes, we sequenced complete mtDNA for both these strains and identified 110 variable positions between F344 and GK sequences. This large amount of variation made it difficult to prioritize functionally important positions. To address this question, we sequenced an additional 11 mitochondrial genomes from various commonly used inbred rat strains including a wild rat from Sweden. Our aim was to identify functionally important sites based on phylogenetic analyses. We used three different tests to find evidence for deviation from neutrality in different parts of the mitochondrial genome. For this purpose we separately analyzed the protein-coding-regions, RNA-coding-regions and the non-coding-regions. Analysis of the protein-coding region provided evidence for selection. We could identify three genes (ND2, ND4, and Cytb) and individual sites within them that were experiencing selection. We

propose that these sites are of functional importance and should be candidates for further investigations. We found additional evidence of selection in the 16S rRNA gene. Despite high substitution rate in the regulatory D-loop as compared to the protein-coding regions we did not find any evidence of selection in the D-loop. However, using comparative analyses, we identified several nucleotide positions within the D-loop that likely play a role in mtDNA transcription and replication.

Understanding the mechanisms of complex polygenic diseases like T2D is as complex as their nature. Cellular processes involved in glucose homeostasis need high and reliable energy supply. Mitochondria are the major energy producers in mammalian cells. Since the majority of proteins active in the mitochondria are encoded by the nuclear genome a tightly regulated and coordinated nuclear-mitochondrial crosstalk needs to be ensured. Perturbation of this communication will attenuate the cellular response to variable energy demands. In this study we have identified nuclear loci which are modulated by the mitochondrial genotype and identified mitochondrial loci that might be of vital importance for its function. The results of our study will provide additional stimulus for further research in understanding nuclear-mitochondrial interactions and their role in metabolic diseases like T2D.

Acknowledgements

I see this day only because of some exceptional people and want to take this opportunity to convey my feelings towards them. First and foremost I would like to thank my family who believed in me and supported me in every walk of my life. This work would never have been possible without the support and encouragement from my beloved and spirited wife, Amruta. My adorable son, Yash, you have always brought fun and laughter to us. I love you both so much ♥.

As I look back into all these years of my existence in Sweden I realize that there are not many places on earth where you come across so many wonderful people. This can happen only in Sweden!

I am deeply grateful to Prof. Holger Luthman for accepting me as his student. It has been a privilege and absolute pleasure to be a part of your team. Your enthusiasm, positive attitude and the never ending quest to learn new things will always inspire me. You have taught me the real meaning of SCIENTIFIC RESEARCH.

I also thank Giancarlo Tonolo for sharing your knowledge and skills and all the support you provided over these years. I will always remember your generosity and cheerful attitude towards everything in life.

Prof. Leif Groop, thank you very much for the encouragement and inspiring comments.

Thank you Charlotte Granhall for teaching me the core practical skills that laid the foundation for this work. I would also like to sincerely thank Agneta Petersson for sharing your skills and wisdom. I admire your enthusiasm and attitude towards acquiring new skills. I thank Inga-Britt Jönsson for making me feel at home right from the day I joined the lab. I also thank Anne Vähäniemi for helping me with the administrative queries.

Many thanks to Hee-Bok Park for sharing your in-depth knowledge and for all those productive scientific and non-scientific discussions. You have always been there to help me. I also thank Ramunas Jokubka for being a wonderful friend.

Special thanks to Lisa Jansson for teaching me sooo many things, especially Swedish! I will miss your delicious cakes. Thank you very much for providing me advice and help whenever I needed it.

Sincere thanks to Nikoo Fathi, Anders Långberg and Elin Hall for being wonderful colleagues and scientists. Wish you good luck for the future. I would also like to thank Karin Berger, Eva Degerman, Anders Rosengren, Erik Renström, Sofia Lagerholm, Kristiana Åkesson, Yeudan Zhou, Ola Hansson for the fruitful collaboration and helping with the experiments in spite of your busy schedules.

Hemang Parikh, you have been a absolutely wonderful friend. Thank you for all the nice moments you shared with me and my family. I wish you the very best in life.

I would also like to thank all my friends and colleagues at CRC for providing unconditional help and support. It has truly been a celestial experience!

References

1. International Diabetes Federation (www.idf.org).
2. Wei, M., Gaskill, S.P., Haffner, S.M. & Stern, M.P. Effects of diabetes and level of glycemia on all-cause and cardiovascular mortality. The San Antonio Heart Study. *Diabetes Care* **21**, 1167-72 (1998).
3. Nichols, G.A., Glauber, H.S. & Brown, J.B. Type 2 diabetes: incremental medical care costs during the 8 years preceding diagnosis. *Diabetes Care* **23**, 1654-9 (2000).
4. ADA. Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **32**, S62-S67 (2009).
5. Pierce, M., Keen, H. & Bradley, C. Risk of diabetes in offspring of parents with non-insulin-dependent diabetes. *Diabet Med* **12**, 6-13 (1995).
6. Tattersal, R.B. & Fajans, S.S. Prevalence of diabetes and glucose intolerance in 199 offspring of thirty-seven conjugal diabetic parents. *Diabetes* **24**, 452-62 (1975).
7. Kaprio, J. et al. Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* **35**, 1060-7 (1992).
8. Newman, B. et al. Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia* **30**, 763-8 (1987).
9. Maassen, J.A., t Hart, L.M. & Ouwens, D.M. Lessons that can be learned from patients with diabetogenic mutations in mitochondrial DNA: implications for common type 2 diabetes. *Curr Opin Clin Nutr Metab Care* **10**, 693-7 (2007).
10. Whittaker, R.G. et al. Prevalence and progression of diabetes in mitochondrial disease. *Diabetologia* **50**, 2085-9 (2007).
11. Alcolado, J.C. & Alcolado, R. Importance of maternal history of non-insulin dependent diabetic patients. *BMJ* **302**, 1178-80 (1991).
12. Alcolado, J.C. et al. Mitochondrial gene defects in patients with NIDDM. *Diabetologia* **37**, 372-6 (1994).
13. Alcolado, J.C. & Thomas, A.W. Maternally inherited diabetes mellitus: the role of mitochondrial DNA defects. *Diabet Med* **12**, 102-8 (1995).
14. Gerbitz, K.D., Gempel, K. & Brdiczka, D. Mitochondria and diabetes. Genetic, biochemical, and clinical implications of the cellular energy circuit. *Diabetes* **45**, 113-26 (1996).
15. Gerbitz, K.D., Paprotta, A., Jaksch, M., Zierz, S. & Drechsel, J. Diabetes mellitus is one of the heterogeneous phenotypic features of a mitochondrial DNA point mutation within the tRNALeu(UUR) gene. *FEBS Lett* **321**, 194-6 (1993).
16. Lee, H.K., Park, K.S., Cho, Y.M., Lee, Y.Y. & Pak, Y.K. Mitochondria-based model for fetal origin of adult disease and insulin resistance. *Ann N Y Acad Sci* **1042**, 1-18 (2005).
17. Mathews, C.E. & Berdanier, C.D. Noninsulin-dependent diabetes mellitus as a mitochondrial genomic disease. *Proc Soc Exp Biol Med* **219**, 97-108 (1998).

18. Wallace, D.C. Diseases of the mitochondrial DNA. *Annu Rev Biochem* **61**, 1175-212 (1992).
19. Bensch, K.G. et al. Selective mtDNA mutation accumulation results in beta-cell apoptosis and diabetes development. *Am J Physiol Endocrinol Metab* **296**, E672-80 (2009).
20. Reardon, W. et al. Diabetes mellitus associated with a pathogenic point mutation in mitochondrial DNA. *Lancet* **340**, 1376-9 (1992).
21. Ristow, M. et al. Frataxin deficiency in pancreatic islets causes diabetes due to loss of beta cell mass. *J Clin Invest* **112**, 527-34 (2003).
22. Fleming, J.C. et al. The gene mutated in thiamine-responsive anaemia with diabetes and deafness (TRMA) encodes a functional thiamine transporter. *Nat Genet* **22**, 305-8 (1999).
23. Mulder, H. & Ling, C. Mitochondrial dysfunction in pancreatic beta-cells in Type 2 diabetes. *Mol Cell Endocrinol* **297**, 34-40 (2009).
24. Maechler, P. & Wollheim, C.B. Mitochondrial function in normal and diabetic beta-cells. *Nature* **414**, 807-12 (2001).
25. Mootha, V.K. et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-73 (2003).
26. Domenech, E., Gomez-Zaera, M. & Nunes, V. WFS1 mutations in Spanish patients with diabetes mellitus and deafness. *Eur J Hum Genet* **10**, 421-6 (2002).
27. Sandhu, M.S. et al. Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* **39**, 951-3 (2007).
28. Labay, V. et al. Mutations in SLC19A2 cause thiamine-responsive megaloblastic anaemia associated with diabetes mellitus and deafness. *Nat Genet* **22**, 300-4 (1999).
29. DiMauro, S. & Schon, E.A. Mitochondrial respiratory-chain diseases. *N Engl J Med* **348**, 2656-68 (2003).
30. Matschinsky, F.M. Banting Lecture 1995. A lesson in metabolic regulation inspired by the glucokinase glucose sensor paradigm. *Diabetes* **45**, 223-41 (1996).
31. Wiederkehr, A. & Wollheim, C.B. Minireview: implication of mitochondria in insulin secretion and action. *Endocrinology* **147**, 2643-9 (2006).
32. Berg, J.M., J.L., Tymoczko, L., Stryer. Biochemistry (5th e.d.). *W.H. Freeman, New York, NY* (2002).
33. Manoli, I. et al. Mitochondria as key components of the stress response. *Trends Endocrinol Metab* **18**, 190-8 (2007).
34. Luft, R. The development of mitochondrial medicine. *Proc Natl Acad Sci U S A* **91**, 8731-8 (1994).
35. Wallace, D.C. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* **39**, 359-407 (2005).

36. Andersson, S.G., Karlberg, O., Canback, B. & Kurland, C.G. On the origin of mitochondria: a genomics perspective. *Philos Trans R Soc Lond B Biol Sci* **358**, 165-77; discussion 177-9 (2003).
37. Margulis, L. Origin of Eukaryotic Cells. *Yale Univ. Press, New Haven, CT* (1970).
38. Blier, P.U., Dufresne, F. & Burton, R.S. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends Genet* **17**, 400-6 (2001).
39. Rand, D.M., Haney, R.A. & Fry, A.J. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol* **19**, 645-53 (2004).
40. Ryan, M.T. & Hoogenraad, N.J. Mitochondrial-nuclear communications. *Annu Rev Biochem* **76**, 701-22 (2007).
41. McCarthy, M.I. Growing evidence for diabetes susceptibility genes from genome scan data. *Curr Diab Rep* **3**, 159-67 (2003).
42. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881-5 (2007).
43. Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-6 (2007).
44. Scott, L.J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341-5 (2007).
45. Shtir, C. et al. Subsets of Finns with high HDL to total cholesterol ratio show evidence for linkage to type 2 diabetes on chromosome 6q. *Hum Hered* **63**, 17-25 (2007).
46. Omori, S. et al. Association of CDKAL1, IGF2BP2, CDKN2A/B, HHEX, SLC30A8, and KCNJ11 with susceptibility to type 2 diabetes in a Japanese population. *Diabetes* **57**, 791-5 (2008).
47. Pascoe, L. et al. Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. *Diabetes* **56**, 3101-4 (2007).
48. Steinthorsdottir, V. et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* **39**, 770-5 (2007).
49. Zeggini, E. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336-41 (2007).
50. Salonen, J.T. et al. Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *Am J Hum Genet* **81**, 338-45 (2007).
51. Yasuda, K. et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* **40**, 1092-7 (2008).
52. Unoki, H. et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* **40**, 1098-102 (2008).

53. Hanson, R.L. et al. A search for variants associated with young-onset type 2 diabetes in American Indians in a 100K genotyping array. *Diabetes* **56**, 3045-52 (2007).
54. Hayes, M.G. et al. Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes* **56**, 3033-44 (2007).
55. Bar-On, H., Ben-Sasson, R., Ziv, E., Arar, N. & Shafrir, E. Irreversibility of nutritionally induced NIDDM in Psammomys obesus is related to beta-cell apoptosis. *Pancreas* **18**, 259-65 (1999).
56. Etgen, G.J. & Oldham, B.A. Profiling of Zucker diabetic fatty rats in their progression to the overt diabetic state. *Metabolism* **49**, 684-8 (2000).
57. Goto, Y., Kakizaki, M. & Masaki, N. Production of spontaneous diabetic rats by repetition of selective breeding. *Tohoku J Exp Med* **119**, 85-90 (1976).
58. Kawano, K. et al. Spontaneous long-term hyperglycemic rat with diabetic complications. Otsuka Long-Evans Tokushima Fatty (OLETF) strain. *Diabetes* **41**, 1422-8 (1992).
59. Nakhooda, A.F., Like, A.A., Chappel, C.I., Wei, C.N. & Marliss, E.B. The spontaneously diabetic Wistar rat (the "BB" rat). Studies prior to and during development of the overt syndrome. *Diabetologia* **14**, 199-207 (1978).
60. Nakhooda, A.F., Wei, C.N., Like, A.A. & Marliss, E.B. The spontaneously diabetic Wistar rat (the "BB" rat): the significance of transient glycosuria. *Diabete Metab* **4**, 255-9 (1978).
61. Pravenec, M. et al. Direct linkage of mitochondrial genome variation to risk factors for type 2 diabetes in conplastic strains. *Genome Res* **17**, 1319-26 (2007).
62. Ziv, E., Shafrir, E., Kalman, R., Galer, S. & Bar-On, H. Changing pattern of prevalence of insulin resistance in Psammomys obesus, a model of nutritionally induced type 2 diabetes. *Metabolism* **48**, 1549-54 (1999).
63. Krinke, G. *The Laboratory Rat*, (Academic Press, 2000).
64. Gibbs, R.A. et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493-521 (2004).
65. Mashimo, T., Voigt, B., Kuramoto, T. & Serikawa, T. Rat Phenome Project: the untapped potential of existing rat strains. *J Appl Physiol* **98**, 371-9 (2005).
66. Saar, K. et al. SNP and haplotype mapping for genetic analysis in the rat. *Nat Genet* **40**, 560-6 (2008).
67. Jansen, R.C. Interval mapping of multiple quantitative trait loci. *Genetics* **135**, 205-11 (1993).
68. Zeng, Z.B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc Natl Acad Sci U S A* **90**, 10972-6 (1993).
69. Hamosh, A., Scott, A.F., Amberger, J., Valle, D. & McKusick, V.A. Online Mendelian Inheritance in Man (OMIM). *Hum Mutat* **15**, 57-61 (2000).

70. Lenffer, J. et al. OMIA (Online Mendelian Inheritance in Animals): an enhanced platform and integration into the Entrez search interface at NCBI. *Nucleic Acids Res* **34**, D599-601 (2006).
71. Safran, M. et al. GeneCards 2002: towards a complete, object-oriented, human gene compendium. *Bioinformatics* **18**, 1542-3 (2002).
72. Hoffmann, R. & Valencia, A. A gene network for navigating the literature. *Nat Genet* **36**, 664 (2004).
73. Jensen, L.J. et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**, D412-6 (2009).
74. Marchler-Bauer, A. et al. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* **37**, D205-10 (2009).
75. Barrett, T. et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* **37**, D885-90 (2009).
76. Castillo-Davis, C.I., Hartl, D.L. & Achaz, G. cis-Regulatory and protein evolution in orthologous and duplicate genes. *Genome Res* **14**, 1530-6 (2004).
77. Brunham, L.R. et al. Accurate Prediction of the Functional Significance of Single Nucleotide Polymorphisms and Mutations in the ABCA1 Gene. *PLoS Genet* **1**, e83 (2005).
78. Thomas, P.D. et al. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research* **13**, 2129-2141 (2003).
79. Mears, J.A. et al. A structural model for the large subunit of the mammalian mitochondrial ribosome. *J Mol Biol* **358**, 193-212 (2006).
80. Cantatore, P., Daddabbo, L., Fracasso, F. & Gadaleta, M.N. Identification by in organello footprinting of protein contact sites and of single-stranded DNA sequences in the regulatory region of rat mitochondrial DNA. Protein binding sites and single-stranded DNA regions in isolated rat liver mitochondria. *J Biol Chem* **270**, 25020-7 (1995).
81. Larizza, A., Pesole, G., Reyes, A., Sbisa, E. & Saccone, C. Lineage specificity of the evolutionary dynamics of the mtDNA D-loop region in rodents. *J Mol Evol* **54**, 145-55 (2002).
82. Roberti, M. et al. Multiple protein-binding sites in the TAS-region of human and rat mitochondrial DNA. *Biochem Biophys Res Commun* **243**, 36-40 (1998).
83. Sbisa, E., Tanzariello, F., Reyes, A., Pesole, G. & Saccone, C. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* **205**, 125-40 (1997).
84. Tipton, C.M. Susruta of India, an unrecognized contributor to the history of exercise physiology. *J Appl Physiol* **104**, 1553-1556 (2008).
85. King, M.P. & Attardi, G. Human cells lacking mtDNA: repopulation with exogenous mitochondria by complementation. *Science* **246**, 500-3 (1989).

86. Forner, F., Foster, L.J., Campanaro, S., Valle, G. & Mann, M. Quantitative proteomic comparison of rat mitochondria from muscle, heart, and liver. *Mol Cell Proteomics* **5**, 608-19 (2006).
87. Ozawa, T., Sako, Y., Sato, M., Kitamura, T. & Umezawa, Y. A genetic approach to identifying mitochondrial proteins. *Nat Biotechnol* **21**, 287-93 (2003).
88. Prokisch, H. et al. Integrative analysis of the mitochondrial proteome in yeast. *PLoS Biol* **2**, e160 (2004).
89. Taylor, S.W. et al. Characterization of the human heart mitochondrial proteome. *Nat Biotechnol* **21**, 281-6 (2003).
90. Calvo, S. et al. Systematic identification of human mitochondrial disease genes through integrative genomics. *Nat Genet* **38**, 576-82 (2006).
91. Steinmetz, L.M. et al. Systematic screen for human disease genes in yeast. *Nat Genet* **31**, 400-4 (2002).
92. Pagliarini, D.J. et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112-23 (2008).
93. Sugiana, C. et al. Mutation of C20orf7 disrupts complex I assembly and causes lethal neonatal mitochondrial disease. *Am J Hum Genet* **83**, 468-78 (2008).
94. Graham, B.H. et al. A mouse model for mitochondrial myopathy and cardiomyopathy resulting from a deficiency in the heart/muscle isoform of the adenine nucleotide translocator. *Nat Genet* **16**, 226-34 (1997).
95. Melov, S. et al. A novel neurological phenotype in mice lacking mitochondrial manganese superoxide dismutase. *Nat Genet* **18**, 159-63 (1998).
96. Larsson, N.G. et al. Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice. *Nat Genet* **18**, 231-6 (1998).
97. Park, C.B. et al. MTERF3 is a negative regulator of mammalian mtDNA transcription. *Cell* **130**, 273-85 (2007).
98. Roubertoux, P.L. et al. Mitochondrial DNA modifies cognition in interaction with the nuclear genome and age in mice. *Nat Genet* **35**, 65-9 (2003).
99. Yu, X. et al. Dissecting the effects of mtDNA variations on complex traits using mouse conplastic strains. *Genome Res* **19**, 159-65 (2009).
100. Rand, D.M., Fry, A. & Sheldahl, L. Nuclear-mitochondrial epistasis and drosophila aging: introgression of Drosophila simulans mtDNA modifies longevity in D. melanogaster nuclear backgrounds. *Genetics* **172**, 329-41 (2006).
101. Johnson, K.R., Zheng, Q.Y., Bykhovskaya, Y., Spirina, O. & Fischel-Ghodsian, N. A nuclear-mitochondrial DNA interaction affecting hearing impairment in mice. *Nat Genet* **27**, 191-4 (2001).
102. Lee, G.H. et al. Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632-5 (1996).

103. Horio, F. et al. The HND mouse, a nonobese model of type 2 diabetes mellitus with impaired insulin secretion. *Eur J Endocrinol* **153**, 971-9 (2005).
104. Nakamura, M. & Yamada, K. Studies on a diabetic (KK) strain of the mouse. *Diabetologia* **3**, 212-21 (1967).
105. Yamada, K., Nakamura, M. & Yamashita, K. Light and electron microscopic studies on the adenohypophysis of a diabetic (KK) strain of the mouse. *Z Zellforsch Mikrosk Anat* **79**, 429-45 (1967).
106. Ueda, H. et al. The NSY mouse: a new animal model of spontaneous NIDDM with moderate obesity. *Diabetologia* **38**, 503-8 (1995).
107. Zhang, Y. et al. Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425-32 (1994).
108. Lagerholm, S. et al. Genetic regulation of bone traits is influenced by sex and reciprocal cross in F(2) progeny from GK and F344 rats. *J Bone Miner Res* **24**, 1066-74 (2009).
109. Gilliam, D.M., Mantle, M.A., Barkhausen, D.A. & Tweden, D.R. Effects of acute prenatal ethanol administration in a reciprocal cross of C57BL/6J and short-sleep mice: maternal effects and nonmaternal factors. *Alcohol Clin Exp Res* **21**, 28-34 (1997).
110. Brown, M.A., Brown, A.H., Jr., Jackson, W.G. & Miesner, J.R. Genotype x environment interactions in Angus, Brahman, and reciprocal cross cows and their calves grazing common bermudagrass and endophyte-infected tall fescue pastures. *J Anim Sci* **75**, 920-5 (1997).
111. Klindt, J. & Maurer, R.R. Reciprocal cross effects on growth hormone and prolactin secretion in cattle: influence of genotype and maternal environment. *J Anim Sci* **62**, 1660-5 (1986).
112. Pollard, K.S. et al. A genome-wide approach to identifying novel-imprinted genes. *Hum Genet* **122**, 625-34 (2008).
113. Smith, R.J., Dean, W., Konfortova, G. & Kelsey, G. Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Res* **13**, 558-69 (2003).
114. Babak, T. et al. Global survey of genomic imprinting by transcriptome sequencing. *Curr Biol* **18**, 1735-41 (2008).
115. Yang, H.H., Hu, Y., Edmonson, M., Buetow, K. & Lee, M.P. Computation method to identify differential allelic gene expression and novel imprinted genes. *Bioinformatics* **19**, 952-5 (2003).
116. Ruf, N. et al. Expression profiling of uniparental mouse embryos is inefficient in identifying novel imprinted genes. *Genomics* **87**, 509-19 (2006).
117. Daura-Oller, E., Cabre, M., Montero, M.A., Paternain, J.L. & Romeu, A. A First-Stage Approximation to Identify New Imprinted Genes through Sequence Analysis of Its Coding Regions. *Comp Funct Genomics*, 549387 (2009).

118. Nabika, T., Kobayashi, Y. & Yamori, Y. Congenic rats for hypertension: how useful are they for the hunting of hypertension genes? *Clin Exp Pharmacol Physiol* **27**, 251-6 (2000).
119. Kreutz, R. & Hubner, N. Congenic rat strains are important tools for the genetic dissection of essential hypertension. *Semin Nephrol* **22**, 135-47 (2002).
120. Kloting, I. et al. Diabetes and hypertension in rodent models. *Ann N Y Acad Sci* **827**, 64-84 (1997).
121. Rogner, U.C. & Avner, P. Congenic mice: cutting tools for complex immune disorders. *Nat Rev Immunol* **3**, 243-52 (2003).
122. Henry, T. & Mohan, C. Systemic lupus erythematosus--recent clues from congenic strains. *Arch Immunol Ther Exp (Warsz)* **53**, 207-12 (2005).
123. Cicila, G.T. & Lee, S.J. Identifying candidate genes for blood pressure quantitative trait loci using differential gene expression and a panel of congenic strains. *Hypertens Res* **21**, 289-96 (1998).
124. Gaspari, M., Larsson, N.G. & Gustafsson, C.M. The transcription machinery in mammalian mitochondria. *Biochim Biophys Acta* **1659**, 148-52 (2004).
125. Handschin, C. et al. Abnormal glucose homeostasis in skeletal muscle-specific PGC-1alpha knockout mice reveals skeletal muscle-pancreatic beta cell crosstalk. *J Clin Invest* **117**, 3463-74 (2007).
126. Kristiansen, O.P. & Mandrup-Poulsen, T. Interleukin-6 and diabetes: the good, the bad, or the indifferent? *Diabetes* **54 Suppl 2**, S114-24 (2005).
127. Pedersen, B.K. et al. Searching for the exercise factor: is IL-6 a candidate? *J Muscle Res Cell Motil* **24**, 113-9 (2003).
128. Larsen, C.M. et al. Interleukin-1-receptor antagonist in type 2 diabetes mellitus. *N Engl J Med* **356**, 1517-26 (2007).
129. Mathew, C.G. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nat Rev Genet* **9**, 9-14 (2008).
130. Goel, A. & Boland, C. Recent insights into the pathogenesis of colorectal cancer. *Current opinion in gastroenterology.* **Publish Ahead of Pri**(2009).
131. Lewinski, M.K. et al. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J Virol* **79**, 6610-9 (2005).
132. Akalin, A. et al. Transcriptional features of genomic regulatory blocks. *Genome Biol* **10**, R38 (2009).
133. Taylor, J. Clues to function in gene deserts. *Trends Biotechnol* **23**, 269-71 (2005).
134. de la Calle-Mustienes, E. et al. A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome Res* **15**, 1061-72 (2005).

135. Ovcharenko, I. et al. Evolution and functional classification of vertebrate gene deserts. *Genome Res* **15**, 137-45 (2005).
136. Goios, A., Pereira, L., Bogue, M., Macaulay, V. & Amorim, A. mtDNA phylogeny and evolution of laboratory mouse strains. *Genome Res.* **17**, 293-298 (2007).
137. Mishmar, D. et al. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* **100**, 171-6 (2003).