The 85th Annual Meeting Invited Lecture

THE CONTRIBUTION OF HOST GENETICS TO TUBERCULOSIS PATHOGENESIS

Erwin SCHURR

Abstract Assessment of the contribution of host genetics to human tuberculosis is a long-standing research challenge. Evidence of genetic factors has come primarily from twin studies and risks to first-degree relatives of cases. In addition, inferences of strong genetic influences have come from anecdotal accounts of socially prominent families, population variation in TB incidence and susceptibility to infection, and secular changes in TB severity, incidence and mortality inferred from historical information of contact between different populations, as well as accidental inoculation of vaccinees with *M. tuberculosis*.

Recently, a major tuberculosis susceptibility locus has been mapped to the long arm of human chromosome. A number of host genetic factors have been directly implicated in tuberculosis susceptibility but strong genetic effects on tuberculosis risk have been difficult to detect both by candidate gene and genome-wide association studies. The reason for our current inability to trace strong genetic effects is unknown. However, a number of possible explanations are supported by direct experimental data. For example, it has been shown that host genetic control of susceptibility is limited to specific host *M. tuberculosis* strain combinations. In addition, it is known that proper inclusion of gene environment interactions is of critical importance for the detection of strong host genetic effects on tuberculosis susceptibility.

By contrast, few genetic studies stratify on *M. tuberculosis* or try to model gene-environment interactions. Until now, most of the human genetics studies in tuberculosis have focused on the identification of genetic variants that impact on progression from infection to disease. There are few studies that aim at the identification of genes that impact on resistance to infection with *M. tuberculosis* or genes that control the extent of anti-mycobacterial immunity. Yet, estimates of heritability for these quantitative traits provide clear evidence for an important role of host genetics in anti-mycobacterial immunity.

Recent work involving scientists from South Africa, France and Canada has focused on the study of innate resistance to infection with *M. tuberculosis*. Employing the tuberculin skin test as a tool to evaluate resistance to infection, a major locus (*TST1*) on chromosomal region 11p14 was identified that T-cell independent resistance to *M. tuberculosis*. In addition, a second major locus (*TST2*), on chromosomal region 5p15 was identified that controls the intensity of T-cell mediated delayed type hypersensitivity (DTH) to tuberculin.

These results pave the way for the understanding of the molecular mechanisms involved in resistance to *M. tuberculosis* infection in endemic areas (*TST1*), and for the identification of critical regulators of T-cell dependent DTH to tuberculin (*TST2*). The finding of a strong host genetic control of anti-mycobacterial immunity raises the questions to what extent host genetics will be a barrier to the development of a universally efficacious tuberculosis vaccine. In fact, epidemiological studies in highly endemic areas and experiments in animal models suggest a strong contribution of host genetic factors to vaccine efficacy making the identification of the corresponding genes one of the new frontiers of mycobacterial research.

I will talk about research that we are doing in my lab, and that research is centered towards the role of host genetic factors in tuberculosis. Tuberculosis (TB) is obviously an infectious disease that is usually but not always caused by *M. tuberculosis*. If there is no *M. tuberculosis*, there is no TB. There is absolutely no question. Therefore, is TB a genetic disease? Many people will say no, because it is caused by *M. tuberculosis* but consider this example.

Phenylketonuria (PKU) is a metabolic disease that is caused by phenylalanine, an amino acid. The incidence of the disease

Departments of Human Genetics and Medicine, McGill University, Montreal, PQ, Canada

Correspondence to : Erwin Schurr, Departments of Human Genetics and Medicine, McGill University, Montreal, PQ, Canada
(E-mail : erwin.schurr@mcgill.ca)
(Received 1 Jul. 2010)

is low among Japanese (1: 125,000) but relatively high among Caucasians (～1: 10,000) and common in the Turkish population (1: 2,600). There is an inability of carriers of the PKU mutation to oxidize phenylalanine and that gives rise to the disease symptoms which are very severe. Indeed, affected patients will die in the absence of medical intervention. It is very clear that if there is no phenylalanine, there is no PKU and prevention of consumption of phenylalanine is the treatment for PKU since patients must consume a diet that does not contain phenylalanine. So, is PKU a genetic disease? You can look in any textbook and the answer will always be yes. So, why the difference between TB, which we don't usually consider a genetic disease and a metabolic disease like phenylketonuria?

I think this goes really back over a century when there was a very emotional and heartfelt discussion about the etiology of infectious diseases, and particularly leprosy and TB. For many centuries, people thought that there was a host genetic contribution. But with the discovery of *M. leprae* and *M. tuberculosis* as the causes of leprosy and TB, respectively, this long-held belief seemed to have been refuted. The conclusion appeared obvious: TB is not a genetic disease; TB a microbiological disease. But people should have listened to the really wise men, such as Louis Pasteur who over 100 years ago already wrote, "it is not the microbe that is transmitted from the parents to the offspring but the predisposition to disease."[1)]

So, if we talk about complex genetic diseases like TB, but this is certainly also true for PKU, we do not talk about an absolute measure of a genetic disease. To reveal the genetic factor an environmental trigger is needed, either phenylalanine in the case of PKU, or in the case of TB, *M. tuberculosis*. There is no question that the cause of TB is *M. tuberculosis*. However, while the bacterium is necessary, it is not sufficient for causing TB disease. I want to give you three historical examples of why we believe the suggestion by Louis Pasteur that there is a genetic predisposition to clinical TB is in fact correct. These are three subjectively selected examples. Though there are many others, I find these particularly instructive: the first is the Lübeck accident, second a consideration of the risk of TB recurrence, and finally twin studies.

The Lübeck accident

What happened in 1929 in the small Northern Germany Hanseatic merchant town of Lübeck? In those days the new vaccine against TB which was *M. bovis* BCG, Bacillus Calmette-Guérin, was distributed by the French researchers to hospitals and laboratories around the world. One of the hospitals that received this new vaccine, that was said to protect from TB, was the hospital in Lübeck. The hospital in Lübeck did not follow the instructions that were sent by the French laboratory although the instructions were very clear. The BCG vaccine strain must not be kept in the same incubator as clinical isolates of *M. tuberculosis*. Unfortunately, that is exactly what happened in the Lübeck hospital. BCG and *M. tuberculosis* were cultivated in the same incubator which resulted in a cross contamination of the vaccine culture with fully virulent *M. tuberculosis*.

At that time, the vaccination was such that you would re-suspend the vaccine strain, i.e., BCG, in the milk of the babies and then the babies were fed with the vaccine in the milk, and that happened three times. So, on three occasions, 251 babies were fed with preparations of vaccine that was contaminated with *M. tuberculosis* and the outcome is shown here (Table 1).[2)] Overall, of the 251 infants, 77 died and 61 had serious disease, but 112, 40%, only had mild symptoms of disease. This becomes even more interesting, if we look at the follow-up study that was done. It became clear that these babies were not all given the same amount of *M. tuberculosis*, but in different vaccine preparations different amounts of *M. tuberculosis* were found resulting in differences in the virulence level of the vaccine preparations. Virulence level 1 had no *M. tuberculosis*; virulence level 4 contained high levels of *M. tuberculosis*.

The interplay between virulence level (i.e. the amount of *M. tuberculosis*) and clinical outcomes is very interesting. With low exposure, meaning a small amount of contamination, the number of deaths was minor, while there were mild symptoms in the majority (80%) of infants. As the exposure intensity increased indicating more *M. tuberculosis* in the milk of the infants, it all switched around. Now, death is the majority outcome while mild symptoms are very rare. Together this is really an impressive example of host-environment interactions. So obviously, if we look at the overall numbers, the differences in clinical outcomes among *M. tuberculosis* infected babies were striking. It seems evident that infants were endowed with variable resistance to *M. tuberculosis*, and because they were newborns, this resistance could not be social factors. It is highly likely that innate genetic factors made some of these babies more susceptible than others. Moreover, resistance/suscep-

Table 1 The Lübeck accident[2)]

Virulence level	Number	Disease severity			
		Death	Serious disease	Mild symptoms	No symptoms
1	1	—	—	—	1
2	93	6 (6.5%)	9 (9.7%)	78 (83.8%)	—
3	83	18 (21.7%)	34 (41.0%)	31 (37.3%)	—
4	74	53 (71.6%)	18 (24.3%)	3 (4.1%)	—
Totals	251	77	61	112	1

tibility are relative and not absolute measures, since clinical outcome was strongly depending on exposure intensity.

The recurrence of tuberculosis

Another example I want to give is the risk of recurrence of TB. I put the very busy table up for the sole purpose of giving you the reference (Table 2).[3] The essence of the data on Table 2 is a study of TB recurrence in successfully treated patients that was conducted in Cape Town, South Africa, by Suzanne Verver and colleagues. Here, *M. tuberculosis* isolates of TB patients were molecularly fingerprinted. Patients then underwent successful treatment as shown by absence of culture positive sputum samples. If patients developed recurrent TB, the *M. tuberculosis* isolate was fingerprinted again and shown to be distinct from the initial *M. tuberculosis* strain.

If the investigators compared the incidence of TB in patients who came back with a second episode of TB (i.e. recurrent TB) with the age-adjusted incidence of patients with a single episode, it turned out that risk of recurrent TB was four times more likely than single episode TB. This to me is an absolute critical observation because many of the TB vaccine endeavors are based on the belief that clinical TB to some extent protects from further episodes of TB. This study shows it is the opposite. TB patients have a four times higher risk of coming back with TB over people who did not have an episode of TB before. This strongly suggests that at least a subset of TB patients is suffering from a (possibly genetic) preponderance of developing TB that cannot be overcome by the presence of acquired immunity.

Twin studies

Let us now move on to more genetic design such as twin studies. Twin studies compare disease concordance rates in dizygous twins and monozygous twins. Monozygous twins, as you know, share 100% of their genetic material, whereas dizygous twins on average share only 50% of their genetic material. Therefore, if we are dealing with a genetic cause, we expect that the concordance rate of disease among monozygous twins will be significantly higher than among dizygous twins. That is exactly what has been found in three twin studies of TB.

Table 3 shows two older studies but very large studies that were done in the 30s and early 40s in Germany[4] and in Upstate New York (Table 3).[5] Also there is a more modern study where the difference was still significant but less pronounced than in the older studies.[6] The reasons for that are not entirely clear. But in all three studies there is absolutely no question that the concordance rate of disease among monozygous twins is significantly higher than among dizygous twins. The conclusion from these studies is very simple: host genetic factors are critical determinants TB susceptibility.

Host genetics of tuberculosis

What are these genetic factors? There comes a little bit of bad news. I don't want to reiterate all the studies that have been done, and rather want to point a little bit more to the future, maybe where we find something. So far, only a very few TB susceptibility genes have been convincingly identified. The emphasis here is on "*convincingly*", because it's very easy to come up with evidence that implicates particular genes and literally we have scores probably over 100 studies that implicate different genes and different variants, but reproducibility is generally very poor. So, "convincingly" we only have a very few genes. Then, the question is "why do we have difficulties

Table 2 Risk of recurrence of TB[3]

Re-infection rate of 2.2/100 PYRS corresponds to 4 times the age-adjusted incidence.

Outcome of first episode with DNA FP	No. patients	PYRS follow-up	No. recurrences	Recurrence rate/ 100 PYRS	No. DNA FP in second episode	No. confirmed reinfections (%)	Confirmed reinfection disease rate/100 PYRS (95% CI)	Likely reinfection disease rate/100 PYRS † (95% CI)
Cure*	358	1,794	48	2.7	21	19 (90)	1.1 (0.7−1.7)	2.4 (1.4−3.8)
TC*	89	466	13	2.8	10	5 (50)	1.1 (0.3−2.5)	1.4 (0.5−3.3)
Default	165	725	47	6.5	37	4 (11)	0.6 (0.2−1.4)	0.7 (0.2−1.6)

Definition of abbreviations: CI = confidence interval; FP = fingerprint; PYRS = person-years; TC = treatment completed.
*For successful treatment (either cure or TC) confirmed reinfections are 24 of 31 (77%), confirmed reinfection disease rate is 1.1 (0.7−1.6) per 100 PYRS, and likely reinfection disease rate is 2.2 (1.6−2.9) per 100 PYRS.
† The likely reinfection disease rate is the recurrence rate multiplied with the proportion confirmed reinfections among recurrences with a DNA FP available.

Table 3 Twin studies

Significant causes of concordance among monozygous twins demonstrates the importance of host genetic factors.

Concordance		Reference
Monozygous twins	Dizygous twins	
65%	25%	Diehl & von Verschuer[4]
62%	13%	Kallmann & Reisner[5]
32%	14%	Comstock[6]

in identifying such host susceptibility genes?" My answer to that is, though other people will give you other answers, that TB pathogenesis is more than one single disease.

I was trained as a physical chemist and in thermodynamic we talk about "functions of state." To put it simple, for functions of state one is not concerned on how one is getting from A to B, you only care about being at B (e.g. suffering from TB), while once you were at A (e.g. being healthy). While we usually treat TB disease as a function of state, i.e. all patients are considered the same, this reasoning is likely wrong for the underlying pathways of TB pathogenesis. In TB and other infectious diseases, it matters a lot how you get from being healthy to becoming a patient. To give you an example. If we ask people from around the world to go to Rome and you then ask people after their arrival about their most memorable experience during their journeys, you will get very different answers. Consider a person from South Africa and who may have had to take a bus all across Africa, and compare to person who flies first-class from New York City to Rome, these travelers will have encountered totally different obstacles. That is certainly true for complex diseases such as TB.

I think we have to pay attention on how we actually get the disease, and an important factor that we have to consider is age. We have to look at host-pathogen interaction, meaning we have to take into account the particular strain of *M. tuberculosis*, we need to consider the clinical stage of the disease, and we have to look at gene-environment interactions. I now give you a few select examples that I believe may explain in part the difficulties in finding TB susceptibility genes.

The importance of age

Why is age important for genetics of TB? I would argue that in TB at a different age you have a different pathogenesis and you have a different disease. From a genetic point of view, I would argue that at different ages, different mechanisms of genetic control are involved. What do I mean by that? After exposure to *M. tuberculosis*, a person can either be resistant or susceptible to infection. In the latter case, an infected person may go on and develop primary TB. This can be considered as clinical TB without latency, which is typical for childhood TB. However a substantial proportion of infected persons enter latent TB infection (LTBI) without any overt clinical symptoms. A subset of persons with LTBI will develop clinical TB later in life and this is commonly referred to as reactivation TB. Reactivation TB may happen within the first couple of years after initial infection or later in life, a situation that you now encounter more frequently in Japan, as I just learnt.

If we take TB mortality as a crude surrogate for TB incidence and then look at recorded age distribution of TB mortality throughout the ages, we can see that age is a very important parameter for the disease incidence. We observe a high mortality of pediatric disease that is generally a disseminated disease, with a drop of mortality throughout childhood, followed by the golden age of TB, which is between roughly, let's say, 10 to 20 years where there is a pronounce drop in incidence. Then the pulmonary TB starts popping up which leads to high incidence of geriatric TB at the old age. Not only would I argue that at different ages these are different diseases, i.e., diseases with different manifestations, but the underlying genetic control in these diseases is also different. If we mix these different forms of TB, we will of course mix the different means of genetic control which will result in loss of power to detect genetic effects.

The menu of genetic risk factors

Now, I'm sorry I have to be a little bit technical. I assume

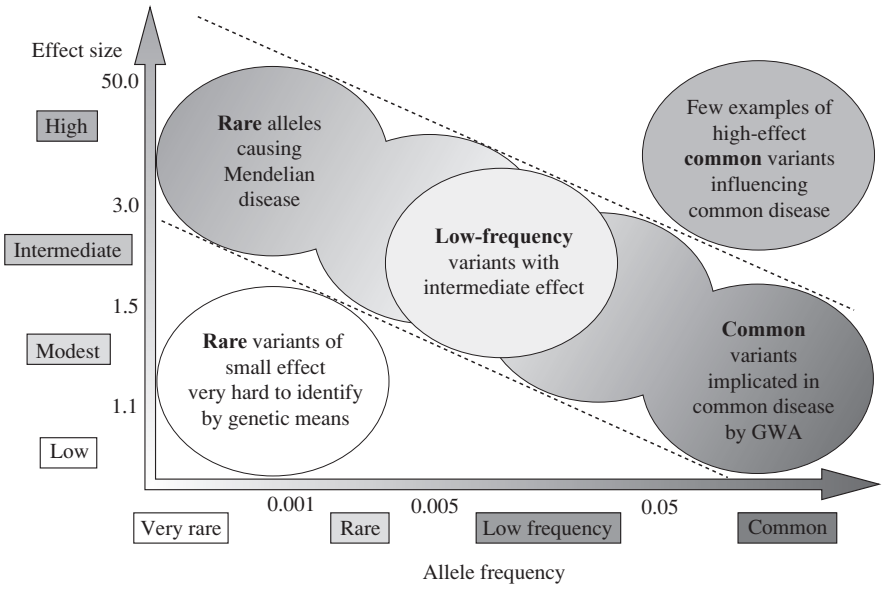


Fig. 1 The menu of genetic risk factors[8)]

you have all heard about genome-wide association studies. The very important assumption that underlies genome wide association studies is that the causative alleles (a particular variant of a given gene) of common diseases such as TB occur at high frequency. To identify these common TB susceptibility variants requires genetic markers that also occur at high frequency. Moreover, if we accept that the causative alleles of TB susceptibility are common these variants must have a weak genetic effect by necessity (Fig. 1). In contrast, if we accept that rare alleles underlie TB susceptibility such rare alleles may have a much stronger effect on TB susceptibility as is shown in Figure 1.[8] Since current genome wide association studies employ genetic markers that display common allele frequencies such studies can only detect common TB susceptibility variants but not rare ones. However, it is possible that the assumption of common TB susceptibility is wrong. A recent genome wide association study of TB in West Africa has indeed failed to implicate any gene in TB susceptibility.[9] These results are unexpected and raise doubts about the general validity of common TB susceptibility alleles. Importantly, the assumption of common susceptibility alleles also underlies most TB case control studies and this may in part explain the poor reproducibility of these studies.

At the other end of this spectrum there are very rare alleles, so rare as maybe only one in 100,000 of people carry actually such an allele. But these alleles may have a very strong effect on TB disease (Fig. 1). My colleagues and I share the belief that these types of rare alleles with strong genetic effects are the genetic cause of pediatric TB. By contrast, common variants with weak genetic effects are thought to play a more important role in adult or geriatric TB disease. The correlation between rare alleles and TB disease is very strong while there is only weak correlation between the common alleles and TB. In other words, in children the genetic control is more straightforward, while in older people genetic control is complicated.[7]

Mendelian predisposition to mycobacterial diseases

Control of TB susceptibility by rare alleles with strong genetic effects is often also called "Mendelian" susceptibility. The proof of principle for this Mendelian type of predisposition in children is given by Mendelian predisposition to mycobacterial infections (Fig. 2).[10)~16)] That is an idea championed by the group of Jean-Laurent Casanova in Paris and now in New York. This group has identified by now a total of 14 different genes that impact in a Mendelian fashion on susceptibility to mycobacteria, usually not *M. tuberculosis* but other mycobacteria with lower virulence. All the genetic defects discovered so far fall within the IL-12−interferon-gamma loop and this has established the critical importance of the IL-12−interferon-gamma loop in immunity to mycobacterial infection.

There are now examples that show that at least certain types of extreme childhood TB are caused by single mutations within some of these genes. This has prompted my colleague Alexandre Alcaïs to do an estimate of how many of these Mendelian-type mutations we need to explain childhood TB (Fig. 3)[7]. This is really a very simple statistical play—a Bayesian approach. What we see here in the formula is the fraction of all childhood TB cases that carry a Mendelian defect. That depends obviously on the frequency of Mendelian defects in children and also on the cumulative frequency of childhood TB. It is a simple formula which makes the outcome all the more striking.

The X-axis of the bottom plain is the frequency of the Mendelian defect, P(Mendel) with a unit of 1 in 10,000. If we have that frequency, i.e., P(Mendel)$=10^{-4}$ and the penetrance, P(cTB\Mendel), of this variant being 90%, then we can explain 45% of all childhood TB cases, i.e., P (Mendel\cTB). Now, 1 in 10,000 means of course not a single gene, it may be 10 genes, or 20 genes, but any way individual genes that carry very rare but very severe mutations. So, we do not need a large number of rare alleles with strong genetic effect and we can already explain a very sizeable proportion of childhood TB. Is this estimate correct or not? We do not know yet. But it is a very attractive working hypothesis, since with the advances in massive parallel sequencing, we will be able to provide the answer to this estimation.

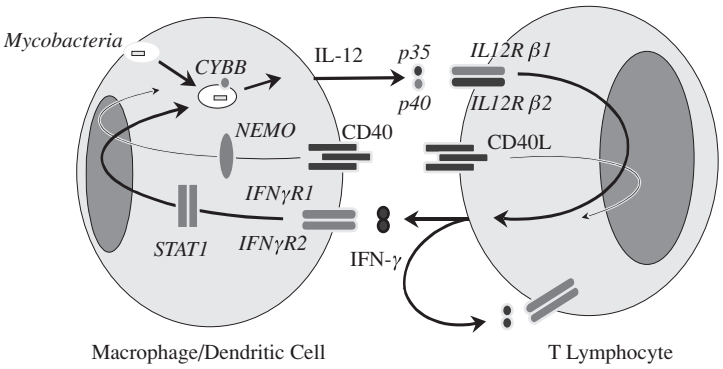
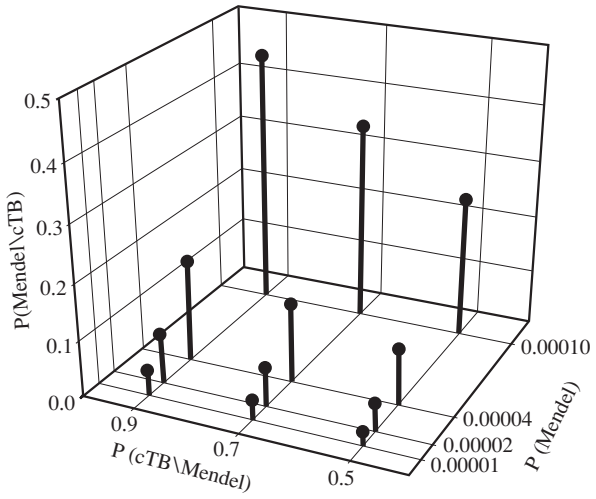


Fig. 2 Mendelian predisposition to mycobacterial infections[10)~16)]



$$P(\text{Mendel}\backslash\text{cTB}) = \frac{P(\text{cTB}\backslash\text{Mendel}) \bullet P(\text{Mendel})}{P(\text{cTB})}$$

$$P(\text{cTB}) \sim 2 \times 10^{-4}.$$

Fig. 3 Mendelian predisposition to disseminated TB in children[7]

Interactions between host genotype and mycobacterial genotype

The next point is host genotype and *M. tuberculosis* genotype interactions. I would like to pick one particular aspect. This is shown here (Fig. 4).[17] Mainly through the work of Maz Divangahi, who was working with Sam Behar and Heinz Remold at Harvard, it has become clear that the balance between prostaglandin E and lipoxin A is critical for the outcome of the infection of a macrophage with *M. tuberculosis*. Lipoxin is an anti-inflammatory mediator and prostaglandin is a pro-inflammatory mediator. If we infect macrophages or dendritic cells with avirulent mycobacteria, the balance goes towards prostaglandin which then triggers TNF and promotes apoptosis. Hence, a prostaglandin dominated response will protect a host from spread of mycobacteria.

If the same type of macrophages are infected with virulent mycobacteria there is no triggering of prostaglandin E synthesis. Rather virulent mycobacteria such as *M. tuberculosis*, induce the synthesis of lipoxin that inhibits COX-2, which is the enzyme catalysing the transformation of arachidonic acid to prostaglandin. Repression of prostaglandin synthesis brings

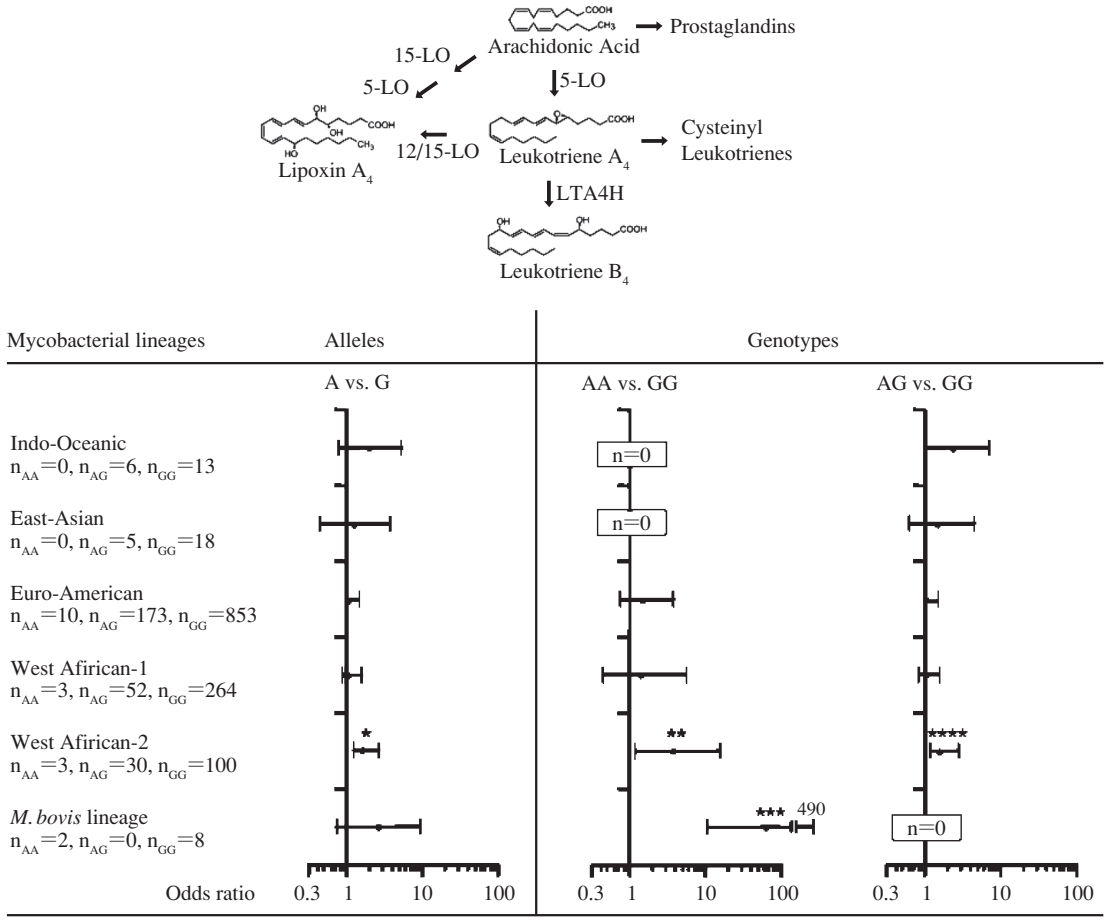


Fig. 4 Strain-specific genetic control[17]
Strain specific effect of ALOX5 (5-lipoxygenase) polymorphisms

down TNF production which in turn promotes necrosis. Necrosis is an unfavorable host reaction, since necrosis results in an uncontrolled release of *M. tuberculosis*. These released bacilli can then be taken up by other cells resulting in a propagation of the infection.

Interactions seen in autophagy

Another aspect of a host defense against mycobacteria is autophagy. We know that avirulent mycobacteria trigger very vigorous autophagy responses whereas infection with virulent mycobacteria inhibits autophagy. So, these are two absolute critical phenomena that decide whether or not an infection with the macrophage can be stopped or whether it will continue uncontrolled, through the balance between prostaglandin and lipoxin and the induction of autophagy.

This provides the background for an example of pathogen specific host genetic control of TB susceptibility. The example comes from the group of Christian Meyer and Rolf Horstmann who studied 5-lipoxygenase polymorphisms in TB susceptibility.[19] Lipoxygenase is one of the enzymes that catalyses the transformation of arachidonic acid to lipoxin as an anti-inflammatory mediator. Mice that are deficient for 5-lipoxygenase are more resistant to *M. tuberculosis* infection. A large case-control study in West Africa analysed the impact of polymorphisms in the *ALOX5* gene which encodes 5-lipoxygenase on risk of TB. Interestingly, an *ALOX5* polymorphism was a significant protective factor for only one substrain of *M. tuberculosis*. Hence, the important aspect here is that for most of the strains there is no impact of this host genotype and if patients are not stratified according to the type of *M. tuberculosis,* this TB protective host genetic factor will be missed. In other words, a host genetic effect is only observed with a particular genotype of *M. tuberculosis*. If hosts are not subdivided according to their *M. tuberculosis* isolates, the effect cannot be shown.

Another example of host-*M. tuberculosis* specific effects in autophagy is a siRNA scan, that was done by the group of Rao in New Delhi, just published in Cell.[20] In essence, these investigators knocked out different genes in THP-1 cells by using pooled siRNAs and then they determined the proliferation of *M. tuberculosis* in THP-1 cells. They came up with 275 genes knocked down by siRNA that impacted on the growth of *M. tuberculosis* in THP-1 cells. Surprisingly, at least to me, is that in 270 out of 275 genes a knockdown of a gene also brought down the number of colony-forming units. For only five genes, the knockdown brought up the number of colony-forming units. So, when we look further into these 270 genes, the knockdown of only 74 genes brings down colony-forming units strain-independently. What that means is that by repeating the same experiment for these 275 genes using seven different clinical isolates it was shown that only 74 of them impacted on all of seven clinical isolates. So, there is a huge discrepancy of genes, with 196 being strain-specific, whereas 74 genes are strain-independent. In addition to a strain-specific effect as we have seen, what is really exciting is that among the 74 genes which are strain-independent, 44 were involved in autophagy whose mechanism is protective for infection with *M. tuberculosis*.

The above finding is in agreement with genetic work again by the group of Meyer and Horstmann[18] who studied immunity-related GTPase *IRGM* which is a regulator of autophagy. They studied the impact of polymorphisms in that gene on genetic susceptibility. They found that there was a very strong effect of this polymorphism on risk of TB, though only in certain strains. When they compared *M. tuberculosis* with *M. africanum*, the entire effect was found in *M. tuberculosis*, while there was absolutely no effect for *M. africanum*. Within the *M. tuberculosis* strains, the genetic effect of *IRGM* was only seen in those *M. tuberculosis* isolates that are incapable of making phenolic glycolipid—all of the Beijing-Asian isolates were insensitive to *IRGM*, whereas the European-North American strains were under control of *IRGM*. This is another very nice and a very strong example of host pathogen genotype interactions. Collectively, these examples show that we need to pay attention to the clinical *M. tuberculosis* isolate when conducting host genetic studies. In most host genetics studies this is not done and may contribute to our difficulty in identifying convincingly TB susceptibility genes.

Gene environment interactions

Regarding gene environment interactions, I quickly want to talk about a study that we did some time ago in Northern Alberta.[21] There was an outbreak of TB and we analyzed that TB outbreak. The entire pedigree involved in this outbreak consisted of 85 individuals, out of whom genotypes were obtained from 65 individuals. The majority of TB cases occurred within 6 months of diagnosis of the index case, and the last TB case was diagnosed 2 years after the index case.

We conducted a parametric linkage analysis where one can put individuals in different liability classes, i.e., groups of varying penetrance of a genetic factor. If a person is in a high liability class, the penetrance is high, and vice versa. Thus, we grouped different people based on their exposure history. Because Canada has a socialized medical system, we know a lot about the medical history of individuals and we were able to assign different people in that family to different liability classes. Employing these liability classes we found a very strong genetic effect on TB susceptibility that is to my knowledge is the strongest genetic effect for non-cosanguineous families ever described in the literature (Table 4).

What is totally striking is that when we conducted the genetic analysis without liability classes, i.e., not taking into account the exposure history of the individuals, then there was zero evidence for a TB susceptibility gene. Remarkably, this is the strongest genetic effect ever described for TB but if we remove information about the exposure history of the family members, this very, very strong effect totally disappears.

Now, this was a unique situation. It was an outbreak. It

Table 4 Genetics of a TB outbreak[21]

Liability class	Penetrance of homozygous			# individuals
	Low risk allele		High risk allele	
	RR＝10	RR＝100		
Previously unexposed	0.085	0.0085	0.85	42
Previously exposed or vaccinated	0.037	0.0037	0.37	11
PPD negative during epidemic	0.010	0.0010	0.10	7
Age ＜2 yrs, ＞65 yrs	0.425	0.2125	0.85	7
Unknown				14

happened at the beginning of the very long Canadian winter. We know the source case. We had good knowledge of the exposure intensities. Clearly, this cannot be a prototype for similar studies. However, as a proof of principle it is a wonderful example in my opinion, since it really makes the point that we need to know more about people than just that they had TB, if we want to do genetic analysis.

Infection and heredity

So far, we have discussed the genetic control of TB disease. But what about infection? I would again like to give you a historical example. The rabbits by Max Lurie are a classic in the genetics of TB because Max Lurie bred two inbred strains of rabbit over many, many years that were either resistant or susceptible to infection with *M. tuberculosis*. This was not an absolute resistance or susceptibility. The resistant rabbits, when infected through injection, actually via the intraperitoneal (IP) route, developed cavitary disease and survived for about 9 months. On the other hand, the susceptible rabbits developed disseminated disease and survived only half of the time of the resistant rabbits. So, to be resistant or susceptible is relative, but it is a very striking difference.[22]

However, Max Lurie was very worried about the effect of the unusual infection procedure, i.e., IP infection. He was wondering if these two strains showed resistance and susceptibility only because of this unusual infection procedure. So, he did an experiment and that experiment lasted 7 years. It is almost never quoted strangely, but just imagine, 7 years—no one can do such studies these days. Even if you wanted to, you could not afford it; you would lose all your grants, and your university would fire you, and that's the end of it, right?

So what was done was that rabbits were infected through the usual route with *M. tuberculosis* and they developed full-blown cavitary disease. They were then put in the center of an arrangement, and around them other rabbits were pushed, either resistant or susceptible types of rabbit, separated by a wire mesh. This provided the setting for a natural aerosol infection in the healthy rabbits. In this arrangement the cavitary rabbits sneeze and cough, and *M. tuberculosis* of course travels through the wire mesh over to the uninfected rabbits and then the uninfected rabbits get the aerosol infection under natural conditions of exposure.

The typical experiment lasted 19 to 24 months and was repeated over the course of, as I said, a total 7 years. In fact it was possible to entirely reproduce the results of the IP infection. Rabbits that belonged to the resistant strains again developed cavitary disease, when infected aerosol in this particular way (there was no aerosolizer at the time), but they survived for 9 months. Rabbits of the susceptible strain developed disseminated disease and they died a lot more quickly.

However, as Werneck-Barosso pointed out, what is virtually never mentioned is the fact that depending on the experiment, 20% to 40% of the rabbits that were exposed up to 24 months to aerosol never developed any symptom of disease and that a vast majority of them tested tuberculin skin negative.[23] Nowadays we would say that these rabbits are resistant to infection, because there is no difference in the time of exposure, and most of infected rabbits showed symptoms after a couple of months. Indeed, uninfected rabbits were kept on average for three times the amount of time of infected rabbits in this arrangement and still did not turn TST positive.

Is there evidence that any person is more or less susceptible to infection? That is of course a lot more difficult to answer. But if we look at cross-sectional studies on age-specific tuberculin positive rates, e.g., one done in Denmark[24] and the other one in India,[25] and if we look at the age distribution of the prevalence of infection, we can note that the prevalence of infection increases with increasing age, but then it plateaus off. There is a difference in the plateau between male and female, but it plateaus for both genders. It plateaus off at around 60% to 70% in Denmark and in India it plateaus off between 50% to 60%. So, it seems that, like in Lurie's rabbits, some people might escape infection, even if they are continuously exposed.

Mycobacterial immunity and LTBI

But now we have to come back and briefly look at mycobacterial immunity. Exposed persons usually become infected via the alveoli, and then intra-alveolar macrophages or interstitial or alveolar macrophages take up *M. tuberculosis*. Then infected phagocytes, usually dendritic cells, migrate into the lymph node and sensitize T-cells. The T-cells in turn migrate back into the lung, where they initiate the formation of a granuloma. This is, of course, all potentiated by different cytokines and chemokines and I do not have to go into that in

detail. The bottom line is that infected macrophages go to the lymph node where sensitization of T-cells happens. Following this sequence of events, we have a typical acquired immunity response that now tries to isolate the infected macrophages in the lung by making a granuloma. We know that this process is effective in more than 90% of the people, since only about 10% of people or less than 10% go actually on from latent TB infection and develop clinical disease.

However, in exposed but infection-resistant persons the migration of the infected phagocyte from the lung to the lymph node never happens presumable because the *M. tuberculosis* insult is dealt with in the lung. In this case, we do not have the sensitization of T-cells and there is no trace of *M. tuberculosis*-specific acquired immunity. Since we cannot isolate *M. tuberculosis* from a subject of LTBI, we need to detect the traces of the presence of *M. tuberculosis* by means of specific acquired immunity. Hence, persons who remain without detectable *M. tuberculosis* specific immunity in the presence of documented exposure may be resistant to *M. tuberculosis*.

There is no gold standard for infection. For many, many years the *in vivo* tuberculin skin test (TST) was used to deduce infection with *M. tuberculosis*. Lately, there are tests that measure the *in vitro* production of antigen-specific interferon gamma by ELISA or the number of antigen-specific interferon gamma producing T-cells with ELISPOT.

We decided to study the genetic control of *M. tuberculosis* infection in Cape Town which has the highest TB incidence in the world. We enrolled 128 nuclear families that had at least two siblings but many of them had more than two. We obtained genomic DNA from the parents and the children, we determined TST reactivities and then we did genotyping. This Figure shows the distribution in the study area of the different families (Fig. 5). The different squares indicate enumerator districts with different incidences. The lowest incidence is around 100 to 200 per 100,000, and the highest is 1,000 to 4,000, that's 1% to 4%. It turned out that in the end it did not matter because infection happened at the level of the community, not at the level of the household or the enumerator district.

I would like to quickly go through over the tuberculin skin test since this was our means of deciding if a person was infected with *M. tuberculosis* and we also used TST reactivity as a measure of the extent of anti-mycobacterial immunity. The TST is done via intradermal injection of PPD and 72 hours later the induration that has developed is being measured. The extent of induration is a read-out of anti-mycobacterial delayed type hypersensitivity. This test has been used for many, many years. Public health has operationally defined certain cut-off points, and based on that a person is considered infected or not-infected. For us it's totally irrelevant, because public health needs simple messages for both educating the public and instructing a myriad of health care workers. So, public health needs operationally defined cut-off points that may not be perfect for any specific condition, —in a basic science study we do not have such restrictions. It's important to realize as scientists that the TST response, i.e., the delayed type immunity, is intrinsically a quantitative measure. And this is a quantitative measure that has very high heritability of 70% to 90%. Heritability determines the proportion of the overall variability that is due to genetic factors. Hence, there is a very strong genetic impact on TST.

As you can see here from our data (Fig. 6), TST reactivity displays crystal clear bimodality. We have a large proportion of individuals with TST＝zero and then we have a beautiful Gaussian distribution that is centered around 14 millimeters induration. Indeed, in our study in Cape Town, the TST offered a better separation of responders and non-responders than the *in vitro* assays.[26)]

There is another way of looking at the same data. Now,

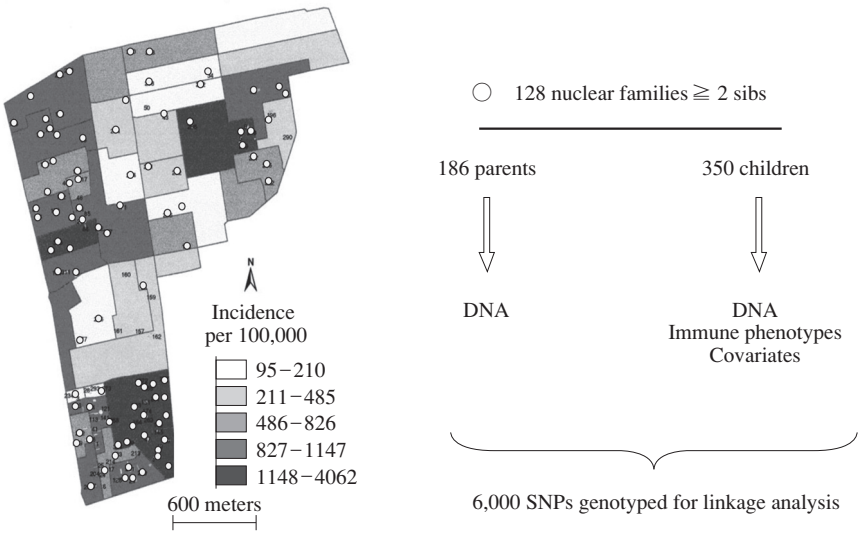


Fig. 5 Study setting: Ravensmead/Uitzag[26)]

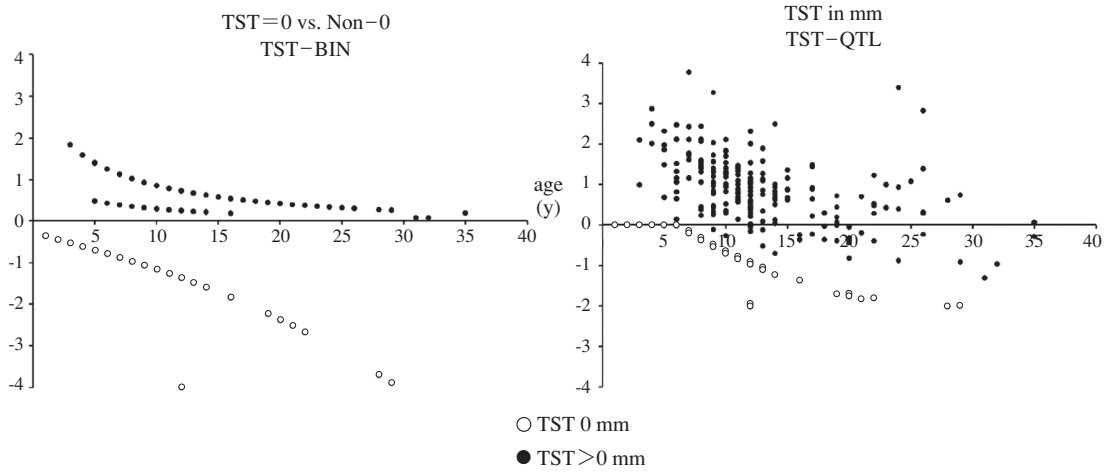


Fig. 6 Two phenotypes — Two linkage analyses[27)]
Vertical axes indicate standardized Pearson residuals (left) and Tobit regression residuals (right).

we plot induration against the age. You can figure that age is important, because the number of zero induration cases decreases as you become older. However, among those with TST＞zero, the extent of induration is largely independent of age.

We wanted to focus on two aspects of the above data. First, we wanted to focus on TST zero versus TST non-zero. This reflects the impressive bimodality that I showed you before. What does that mean? Are there genes that impact on that bimodality? To test for this we performed a logistic regression analysis in the context of covariates including age, gender, and previous TB. In its most basic way, a logistic regression deals with plus or minus outcomes. However, the incorporation of covariates has an impact and that is why the TST phenotypes now show a more continuous distribution (Fig. 6, left). Simplistically, if we observe a person aged 2 years with a TST of zero, it is as expected. If we observe a person who is 45 years old and has a TST zero, now this is something more surprising. Likewise, if we observe a child that is 2 or 3 years old with a very well pronounced TST reactivity, that is not what you expect, so we give more weight to such unexpected observations.

We are also interested in the extent of reactivity, as shown in here (Fig. 6, right). This was done by employing a so-called Tobit regression. You can consider this a form of linear regression that most of you are familiar with, except it is a linear regression that is left-centered because we cannot have values less than zero. In this way, we defined two phenotypes: TST＝zero vs TST＞zero and extent of TST reactivity.

What did we find when we looked for being TST＝zero (i.e. TST negative) vs TST＞zero (i.e. TST positive)? We did a linkage analysis and we found a highly significant hit on chromosome region 11p14, i.e. a major locus that determines either being negative or positive for TST is localized on chromosome 11p14. This locus was called *TST1*. So, this is a gene that determines whether your innate immunity can deal with *M. tuberculosis* that you breathe into your lungs. This is the true TB-infection resistance locus.

When we do the same analysis for extent of TST, we found a major hit on chromosome region 5p15, which signifies that there is a locus that impacts on the extent of anti-mycobacterial immune reactivity that localizes to that chromosomal region. We think that this is some kind of immunoregulatory gene that we were mapping to chromosome 5p15. This locus was called *TST2*. I also would suggest to you that probably these two genes act sequentially. We have first the *TST1*, which is on 11p14, which impacts on staying zero, as a resistance locus, and once you are positive, then *TST2* kicks in to determines how strongly you react. We actually have some additional data to show that this interpretation is correct.[27)]

Future Directions

In the next step we initiated the search for the genes that underlie the two loci mapped in the linkage study and this work is still ongoing. I would like to show you how we approach the problem. The example is a high-density association screen of the *TST2* region. That's the gene that impacts on the extent of reactivity. The approach is that closely spaced genetic markers are genotyped and analysed for association with extent of TST reactivity. The evidence of association of genetic marker with TST reactivity is expressed as a p-value. Small p-values indicate strong association between genetic marker and TST reactivity. In the next figure you see p-values of individual genetic markers that are plotted versus the location of the marker on the chromosome (Fig. 7). Please note that here we use the negative log of p-values, i.e. the smaller the p-value (＝the stronger the association) the higher on the y-axis a genetic marker shows up. Hence, genetic markers in the upper region of the plot display strongly significant evidence that the genetic marker is associated with TST reactivity. As you can see, however, the vast majority of the markers does not display strong evidence for association with TST reactivity.

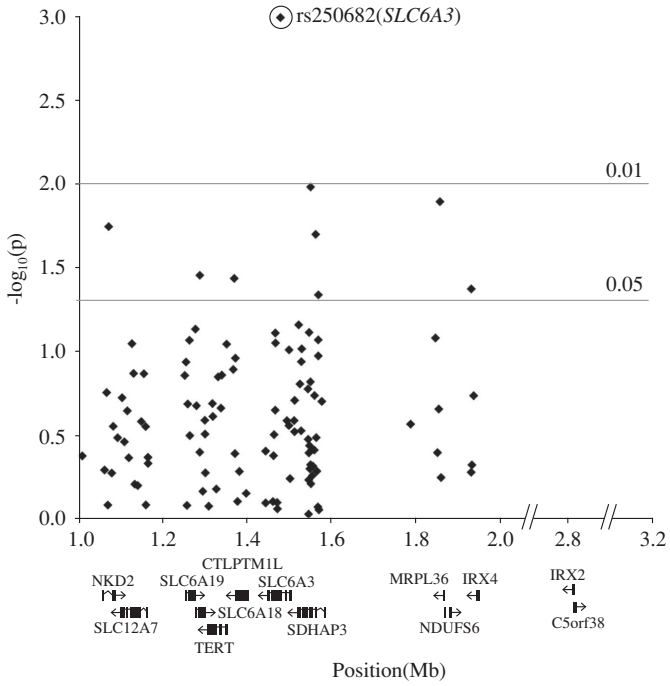


Fig. 7 *TST2* region association scan[27)]

The only exception is a single marker that is located within a gene which is called *SLC6A3*. What is *SLC6A3*? It's a neurotransmitter transporter. It transports dopamine. So, it is the principal regulator of dopaminergic transmission. The gene has been implicated in attention deficits, schizophrenia, substance abuse, Parkinson's disease, and a whole slew of other neurological diseases.

Our first reaction was "Oops!", not what you want to hear. However, when we went further into the literature, we actually found the following study. Knockout mice for *SLC6A3* had been injected IP with ovalbumin twice and these injections were followed 3 weeks later with an ovalbumin challenge injection in the footpad. Following the challenge injection, the footpad swelling was used as a measure of delayed type hypersensitivity. When the extent of delayed type hypersensitivity in the knockout mice was compared with the one displayed by wild-type mice, the wild-type mice had a much more pronounced immune reactivity as compared to the knockout mice.[28)] That, of course, is very interesting in the context of a gene that we have shown to impact on the overall extent of immune reactivity.

The question now is; "Is it true?" I cannot tell you that for sure, because these experiments are in progress, but hopefully soon we will have definite answers for the identity of both *TST1* and *TST2*.

So, obviously there is a lot of work that has gone in here. I would like to point out the people in my lab who were mainly involved. These are C. Gallant, who is now a postdoc in Sweden, Leah Simkin, and Marianna Orlova, a postdoc from Russia in the lab, and Celia Greenwood and Mary Fujiwara who did the linkage study in Northern Alberta. Since many years, we have worked with the group of Laurent Abel, Alexandre Alcaïs and particularly Aurelie Cobat. Aurelie did a lot of the study in Cape Town. In Cape Town we worked with people from Stellenbosch University, Eileen Hoal, Nulda Beyers, Paul van Helden. Willem Hanekom is in immunology from University of Cape Town. Jeane Hughes and Brian Eley are both from University of Cape Town, and Mark Doherty is from Copenhagen who was also involved. Thank you for your attention.

Acknowledgements

INSERM U550, Paris: A. Alcaïs, A. Cobat, L. Abel, JL Casanova
Stellenbosch University, Tygerberg, SA: E. Hoal, N. Beyers, P. van Helden
University of Cape Town, SA: W. Hanekom, J. Hughes, B. Eley
Statens Serum Institute, Copenhagen, Denmark: M. Doherty
Schurr's Lab: S. Malik, C. Gallant, T. Di Pietrantonio, M. Girard, J L. Simkin, M. Orlova, S. Ryoo, CM Greenwood
Canadian Institutes of Health Research (CIHR), Canadian Genetic Diseases Network (CGDN), Sequella/AERAS Global TB Vaccine Foundation, Howard Hughes Medical Institute (HHMI)

References

1) Louis Pasteur; In 'Etudes sur la ma ladie des vers à soie. La pébrine et la flacherie (tome I)'. Gauthier-Villars eds., 1870.

2) Moegling A: Die Epidemiologie der Lübecker Säuglingstuberkulose, Arbeiten a. d. Reichsges-Amt. 1931 ; 69 : 1−24.
3) Verver S, Warren RM, Beyers N, et al.: Rate of reinfection *tuberculosis* after successful treatment is higher than rate of new tuberculosis. Am J Respir Crit Care Med. 2005 ; 171 : 1430−5.
4) Diehl K, von Verschuer O: Der Erbeinfluss bei der Tuberculose. Jena, Germany: Gustav Fisher; 1933. (Diehl and von Verschuer Beitr. Klin Kunsch. 1936 ; 92 : 275).
5) Kallmann FJ, Reisner D: Twin studies on the significance of genetic factors in tuberculosis. Am Rev Tuberc. 1943 ; 47 : 549−574.
6) Comstock GW: Tuberculosis in twins: a re-analysis of the Prophit survey. Am Rev Respir Dis. 1978 ; 117 : 621−4.
7) Alcaïs A, Quintana-Murci L, Thaler DS, et al.: Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? Ann N Y Acad Sci. 2010 ; 1214 : 18−33.
8) Manolio TA, Collins FS, Cox NJ, et al.: Finding the missing heritability of complex diseases. Nature. 2009 ; 461 : 747−753.
9) Thye T, Vannberg FO, Wong SH, et al.: Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet. 2010 ; 42 : 739−41. Epub 2010 Aug 8.
10) Jouanguy E, Altare F, Lamhamedi S, et al.: Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guérin infection. N Engl J Med. 1996 ; 335 : 1956−61.
11) Altare F, Durandy A, Lammas D, et al.: Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. Science. 1998 ; 280 : 1432−5.
12) Dupuis S, Jouanguy E, Al-Hajjar S, et al.: Impaired response to interferon-alpha/beta and lethal viral disease in human STAT1 deficiency. Nat Genet. 2003 ; 33 : 388−91. Epub 2003 Feb 18.
13) Dupuis S, Dargemont C, Fieschi C, et al.: Impairment of mycobacterial but not viral immunity by a germline human STAT1 mutation. Science. 2001 ; 293 : 300−3.
14) Döffinger R, Helbert MR, Barcenas-Morales G, et al.: Autoantibodies to interferon-gamma in a patient with selective susceptibility to mycobacterial infection and organ-specific autoimmunity. Clin Infect Dis. 2004 ; 38 : e10−4. Epub 2003 Dec 4.
15) Filipe Santos O, Bustamante J, Haverkamp M, et al.: X-linked susceptibility to mycobacteria is caused by mutations in the NEMO leucin zipper domain that impair CD40-dependent IL-12 production. J Exp Med. 2006. 203 ; 1745−1759.
16) von Bernuth H, Picard C, Jin Z, et al.: Pyogenic bacterial infections in humans with MyD88 deficiency. Science. 2008 ; 321 : 691−6.
17) Divangahi M, Desjardins D, Nunes-Alves C, et al.: Eicosanoid pathways regulate adaptive immunity to *Mycobacterium tuberculosis*. Nat Immunol. 2010 ; 11 : 751−8. Epub 2010 Jul 11.
18) Intemann CD, Thye T, Niemann S, et al.: Autophagy gene variant IRGM-261T contributes to protection from tuberculosis caused by *Mycobacterium tuberculosis* but not by *M. africanum* strains. PLoS Pathog. 2009 ; 5 : e1000577. Epub 2009 Sep 11.
19) Herb F, Thye T, Niemann S, et al.: ALOX5 variants associated with susceptibility to human pulmonary tuberculosis. Hum Mol Genet. 2008 ; 17 : 1052−60.
20) Kumar D, Nath L, Varshney A, et al.: Genome-wide analysis of the host intracellular network that regulates survival of *Mycobacterium tuberculosis*. Cell. 2010 ; 140 : 731−43.
21) Greenwood CM, Fujiwara TM, Boothroyd LJ, et al.: Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. Am J Hum Genet. 2000 ; 67 : 405−16.
22) Lurie MB: Heredity, constitution and tuberculosis: an experimental study. Am Rev Tuberc. 1941 ; 44 (suppl to No.3) : 1−125.
23) Werneck-Barroso E: Innate resistance to tuberculosis: revisiting Max Lurie genetic experiments in rabbits. Int J Tuberc Lung Dis. 1999 ; 3 :166−68.
24) Groth-Petersen, Knudsen J, Wilbek E: Epidemiological basis of tuberculosis eradication in an advanced country. Bull World Health Organ. 1959 ; 21 : 5−49.
25) National Institute of Tuberculosis, Bangalore: Tuberculosis in rural population of South India: a five-year epidemiological study. Bull WHO. 1959 ; 151 : 473−88.
26) Gallant CJ, Cobat A, Simkin L, et al.: Tuberculin skin test and in vitro assays provide complementary measures of antimycobacterial immunity in children and adolescents. Chest. 2010 ; 137 : 1071−7. Epub 2009 Dec 29.
27) Cobat A, Gallant CJ, Simkin L, et al.: Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. J Exp Med. 2009 ; 206 : 2583−91. Epub 2009 Nov.
28) Kavelaars A, Cobelens PM, Teunis MA, et al.: Changes in innate and acquired immune responses in mice with targeted deletion of the dopamine transporter gene. J Neuroimmunol. 2005 ; 161 : 162−8.