



# An SNP selection strategy identified IL-22 associating with susceptibility to tuberculosis in Chinese

Guoliang Zhang<sup>1\*</sup>, Xinchun Chen<sup>1\*</sup>, Long Chan<sup>2,4</sup>, Mingxia Zhang<sup>1</sup>, Baohua Zhu<sup>3</sup>, Lantian Wang<sup>1</sup>, Xiuyun Zhu<sup>1</sup>, Jieyun Zhang<sup>1</sup>, Boping Zhou<sup>1</sup> & Junwen Wang<sup>4</sup>

<sup>1</sup>Shenzhen–Hong Kong Institute of Infectious Disease, Shenzhen Third People's Hospital, Guangdong Medical College, Shenzhen, 518020, China, <sup>2</sup>Department of Mathematics, Carleton College, 100 South College Str. Northfield, MN, USA, <sup>3</sup>Surgical Department, The First Affiliated Clinical College, Guangdong Medical College, Zhanjiang 524023, China, <sup>4</sup>Department of Biochemistry, LKS Faculty of Medicine, The University of Hong Kong, 21 Sassoon Rd, Pokfulam, Hong Kong SAR, China.

Received  
9 May 2011

Accepted  
2 June 2011

Published  
21 June 2011

Correspondence and requests for materials should be addressed to J.W. (junwen@uw.edu)

\* these authors equally contributed to this manuscript

Recent studies showed that IL-22 plays a protective role in the host defense. However, the contribution of polymorphisms of the IL-22 gene to human TB susceptibility remains untested. We have designed a computational approach to select functional SNPs in the IL-22 gene and genotyped them in a two-stage case-control study in Chinese (479 cases and 358 controls). We found that rs2227473, an SNP in the promoter region of IL-22, is associated with TB susceptibility at both stages of our study. The SNP shows associations with p-values of 0.028 and 0.034 respectively, and a combined p-value of 0.0086, with odds ratio at 0.65 (95% confidence interval 0.45–0.90). We further validated the association with an independent cohort (413 cases and 241 controls in Chinese). Our functional studies showed that patients with A allele have significantly higher IL-22 expression than those without A allele under both non-specific and specific stimulations.

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* infection, is an infectious disease that is reemerging as a major killer of humans. It has been estimated that one third of the world's population are infected by TB, with more than nine million becoming sick each year with “active” TB, which can be spread to healthy people<sup>1</sup>. Most people infected with *Mycobacterium tuberculosis* do not develop active diseases, a status called “latent tuberculosis infection (LTBI)”<sup>2</sup>; only 5–10% eventually develop active tuberculosis diseases, indicating that host response plays a critical role in containment of TB infection. Indeed, both animal and clinical studies have demonstrated that cellular immunity, particularly the CD4<sup>+</sup> T cell immunity, plays indispensable roles in protective immune responses against TB infection<sup>2</sup>.

So far, a large number of polymorphisms in the human leucocyte antigens (HLA) have been discovered to be associated with TB susceptibility in a wide range of ethnic populations<sup>3</sup>. The variations in cytokines and their receptors have also been shown to influence susceptibility to TB. For example, A874T polymorphism in the first intron of interferon (IFN)- $\gamma$  gene was reported to be associated with TB susceptibility in Sicilians<sup>4</sup>, South Africans<sup>5</sup>, and Chinese<sup>6</sup>. Other cytokines, such as IL-1, IL-2, IL-4, IL6, IL-10, IL-12, TNF- $\alpha$ , and TGF- $\beta$ , have also been associated with TB<sup>3</sup>. A recent genome wide association study identified a single nucleotide polymorphism (SNP) in the gene desert region of 18q11.2 that was associated with tuberculosis with p-value of  $6.8 \times 10^{-9}$  in Ghana and Gambia populations<sup>7</sup>.

It was recently reported that IL-17 and IL-22, produced by T helper type 17 (Th17) cells and NK cells, play important roles in TB protective immune response<sup>8,9</sup>. In vitro experiments showed that IL-22 can directly affect human macrophages, improving cells' protective immunity against TB infection<sup>10</sup>. IL-22 has also been shown to induce host defense genes, and thus is crucial for controlling the gram-negative pulmonary pathogen, *Klebsiella pneumoniae*<sup>11</sup>. Knock-out study observed that IL-22 null mice showed increased intestinal epithelial damage, systemic bacterial burden and mortality after *Citrobacter rodentium* infection<sup>12</sup>. It was also reported that IL-22 can inhibit intracellular mycobacterial growth by enhancing phagolysosomal fusion, which contributes to the immune defenses of NK cells against *M. tuberculosis*<sup>10</sup>.

We thus hypothesize that the genetic variations in the IL-22 gene among humans play an important role in TB susceptibility. Based on our previous study<sup>13</sup>, we are interested in the SNPs in the IL-22 gene region that could affect TB susceptibility, particularly the SNPs that could affect the regulation of this gene. To address this



**Table 1** | The association of TB susceptibility and the rs2227473 marker in the promoter region of IL-22 in multiple stages of association studies in Chinese.

| Cohort              | Stage                | Sample counts |          | MAF   |          |              | 95% CI* |       |       | P-value <sup>†</sup> |
|---------------------|----------------------|---------------|----------|-------|----------|--------------|---------|-------|-------|----------------------|
|                     |                      | Cases         | Controls | Cases | Controls | Minor allele | OR      | LCL   | UCL   |                      |
| <b>Experimental</b> | <b>I</b>             | 48            | 48       | 0.021 | 0.103    | A            | 0.188   | 0.037 | 0.967 | 0.0275               |
|                     | <b>II</b>            | 431           | 310      | 0.107 | 0.148    | A            | 0.694   | 0.494 | 0.975 | 0.0343               |
|                     | <b>I+II Combined</b> | 479           | 358      | 0.098 | 0.144    | A            | 0.653   | 0.449 | 0.896 | 0.0086               |
| <b>Validation</b>   | -                    | 413           | 241      | 0.086 | 0.114    | A            | 0.702   | 0.484 | 1.107 | 0.061                |
| <b>E+V Combined</b> | -                    | 892           | 599      | 0.092 | 0.132    | A            | 0.663   | 0.518 | 0.847 | 0.001                |

\*95% CI; lower and upper confidence limits. <sup>†</sup>Allelic p-value.

hypothesis, we carried out extensive genotyping experiments in the gene and evaluated their relationship with TB susceptibility.

We adopted a two-stage approach to genotype the selected SNPs for the experimental cohort, and further validated our discovery in a validation cohort. In the first stage, we scanned a relatively large number of SNPs in a limited number of samples to identify the markers with potential interest. We then selected a small number of SNPs that showed promise, and typed them in a large number of independent samples. The two-stage scan has been adopted in many genome-wide association studies and has been shown to be effective. In this study, during the first stage, we typed 48 cases and 48 controls, which were randomly selected from the pools. We then selected 6 SNPs that met our selection criteria and typed the remaining 431 cases and 310 controls. We finally genotyped one SNP (rs2227473) in the independent validation cohort (413 cases and 241 controls) to further confirm our study. The biological functions (target gene IL-22's expression) of different genotypes under specific and non-specific stimulations were investigated and compared.

## Results

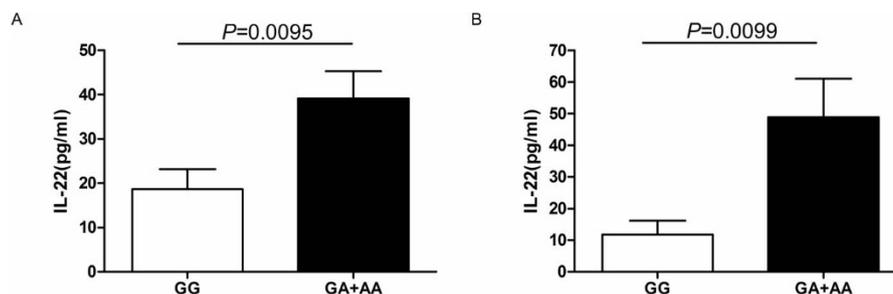
**Reduced susceptibility of the rs2227473A allele in patients with TB in stage I data.** In stage I of 48 cases and 48 controls, the rs2227473 marker, which is located in -1756 bp upstream of the IL-22 gene transcription start site (TSS), was in Hardy-Weinberg equilibrium. The minor allele frequency of 0.103 in the control is comparable to the annotations in the HapMap (0.102 for CHB+JPT, 0.120 for YRI, and 0.208 for CEU). The 'A' allele is associated with decreased risk of TB susceptibility at p-value of 0.0275, with odds ratio (OR) of 0.188 (confidence interval (CI) = 0.037–0.967), as shown in Table 1. The haplotype analysis was further performed in the surrounding region of this SNP. We found that the haplotype TCATGA—which is made of SNPs rs2227485, rs2227484, rs2227483, rs2227478, rs2227473, and rs2227472—shows a trend of association with decreased risk of TB susceptibility at p-value of 4.89E-5 (Supplementary Table S1). We therefore selected these 6 SNPs for genotyping in the second stage.

**Reduced susceptibility of the rs2227473A allele in patients with TB in stage II and pooled data.** The association of rs2227473 was confirmed in the stage II experimental dataset with 431 cases and 310 controls. As shown in Table 1, the association is significant at a p-value of 0.0343, with OR = 0.694 (95% CI = 0.494–0.975). Analysis on the pooled data (479 cases and 358 controls) showed even more significant association at a p-value of 0.0086, with OR=0.653 (95% CI=0.449–0.896). This SNP was predicted to alter the protein-DNA interactions among several transcription factors (TFs), including three TFs from the UniPROBE database, and two TFs from the Jasp database (Supplementary Table S2).

**Marginal association between rs2227473 and TB in validation cohort.** The association was further validated by an independent cohort with 413 cases and 241 controls, which were collected after Feb., 2010 (Table 1). The association was marginally significant in this dataset, with p-value of 0.061, OR=0.702 (95% CI=0.484–1.107). However, when we combined the experimental and validation cohorts, the association was very significant at a p-value of 0.001 (OR=0.663, with 95% CI=0.518–0.847).

**Effects of rs2227473 genotypes on IL-22's protein expression.** To test whether different genotypes of the rs2227473 affect IL-22's expression levels, we stimulated patients' peripheral blood mononuclear cells (PBMCs) with both anti-CD3/anti-CD28 antigens nonspecific and Mtb antigen-specific stimulations. Patients with A allele (GA+AA, n=29) at rs2227473 had significantly higher IL-22 protein productions than those without A allele (GG, n=29) under both non-specific (p value =0.0095) and specific stimulations (p value = 0.0099, Figure 1).

**Other polymorphisms in the IL-22 promoter region that reduce the susceptibility of TB in stage II and pooled data.** Among the other five SNPs we genotyped in stage II of the experimental cohort, four of them showed significant association with TB susceptibility (Table 2). The locations of the four SNPs to the IL-22 TSS are rs2227472 (-1851bp), rs2227478 (-1340bp), rs2227483 (-894bp),



**Figure 1** | IL-22 production by PBMCs from individuals with polymorphism in IL-22 (rs2227473). Subjects were grouped for SNP (rs2227473) in IL-22 as genotype with (GA+AA, n=29) or without (GG, n=29) A allele. Frozen PBMCs were thawed and cultured for 96 hours in the presence of anti-CD3/anti-CD28 (1µg/ml each), or killed Mtb. (20µg/ml). Supernatant was collected and the concentration of IL-22 was determined by ELISA. IL-22 protein production by PBMC in response to (A) anti-CD3/anti-CD28 and (B) Mtb. Data are represented as mean ± SEM. P values are indicated.



**Table 2 |** The association of TB susceptibility and polymorphism of four SNPs in the promoter region of IL-22 in the second stage and two stages pooled studies in Chinese.

| Stage                | Sample counts |          | SNP id    | MAF   |          |     | Minor allele | OR    | 95% CI |        | P-value |
|----------------------|---------------|----------|-----------|-------|----------|-----|--------------|-------|--------|--------|---------|
|                      | Cases         | Controls |           | Cases | Controls | LCL |              |       | UCL    |        |         |
| <b>II</b>            | 431           | 310      | rs2227472 | 0.464 | 0.526    | A   | 0.778        | 0.621 | 0.974  | 0.0285 |         |
|                      |               |          | rs2227478 | 0.148 | 0.189    | C   | 0.746        | 0.564 | 0.988  | 0.0405 |         |
|                      |               |          | rs2227483 | 0.458 | 0.534    | A   | 0.739        | 0.573 | 0.955  | 0.0205 |         |
|                      |               |          | rs2227485 | 0.467 | 0.523    | C   | 0.798        | 0.641 | 0.993  | 0.0432 |         |
| <b>I+II (pooled)</b> | 479           | 358      | rs2227472 | 0.465 | 0.522    | A   | 0.796        | 0.646 | 0.981  | 0.0323 |         |
|                      |               |          | rs2227478 | 0.148 | 0.191    | C   | 0.769        | 0.592 | 0.957  | 0.0220 |         |
|                      |               |          | rs2227483 | 0.46  | 0.528    | T   | 0.762        | 0.604 | 0.962  | 0.0223 |         |
|                      |               |          | rs2227485 | 0.466 | 0.517    | C   | 0.816        | 0.665 | 1.000  | 0.0499 |         |

and rs2227485 (-431bp). SNP rs2227472 is 95 bp upstream of our major SNP, rs2227473. Its A allele is a major allele in controls, but a minor allele in cases in both stage II and pooled data. Similar patterns were observed for two other SNPs, rs2227478 and rs2227485, which are 416 bp, and 1325 bp downstream of rs2227473, respectively. All the four SNPs have p-values in the range of 0.021–0.044, with OR range from 0.74–0.80 (Table 2) for the stage II data. Similar results were observed for the pooled data.

**Haplotypes in the IL-22 promoter region that affect the susceptibility of TB in pooled data.** We observed very high linkage disequilibrium among the five SNPs in the control population of the pooled data (Figure 2). We thus performed the association study between the haplotypes of these five SNPs and TB susceptibility. As shown in Table 3, haplotype CTTAA (SNPs in the order of rs2227485, rs2227483, rs2227478, rs2227473, and rs2227472) shows very strong association with decreased TB susceptibility, at p-value of 2.12E-6, with OR at 0.04 (95% CI=0.01–0.35). On the other hand, the haplotype TATGG, which consists of all major alleles, shows significant association with increased TB susceptibility at p-value of 0.01, with OR at 1.49 (95% CI=1.05–2.13). In the pooled data of the two stages, the p-value is slightly more significant for rs2227478, and a bit less significant for rs2227472, when comparing with stage II alone. The

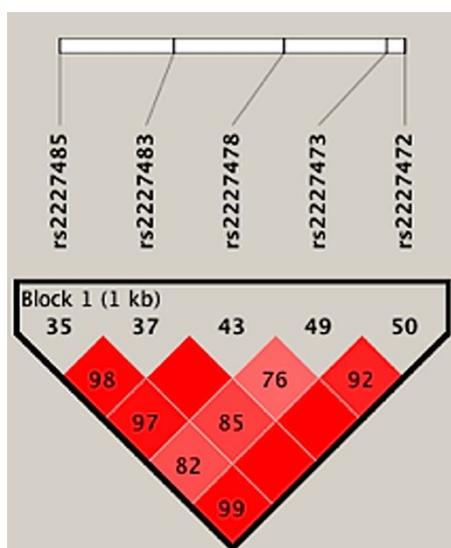
linkage disequilibriums (LDs) of the five SNPs are given in Figure 2, showing moderate LDs.

**The association between SNPs in the IL-22 promoter region and TB susceptibility in pooled data stratified by age at diagnosis ( $\leq 25$  and  $>25$ ).** We further stratified our pooled data according to the age when the patients were admitted to the hospital for the experimental cohort. As shown in Table 4, out of five SNPs showing significant association in the pooled data, three SNPs (rs2227472, rs2227483, and rs2227485) showed more significant association in younger patients but not in elder patients, and two other SNPs (rs2227473 and rs2227478) showed significant association in elder patients but not in younger patients. Notably SNP rs2227485, which is marginally significant in the pooled data (p-value = 0.049), showed very significant association with decreased TB susceptibility in younger patients (age  $\leq 25$  years old) at p-value of 5.2E-5, with OR = 0.496 (95% CI=0.303–0.663). But for the older group (age  $>25$ ), this SNP did not show any association with TB susceptibility. A similar situation is true for SNP rs2227472.

## Discussion

IL-22 has been reported to play a role in host defense against bacterial pathogens<sup>11</sup>. It is involved in host defense at environmental interfaces, such as mucosal surfaces of the airways and gastrointestinal tract. IL-22 can activate STAT3, which is a transcription activator that mediates the expression of a variety of genes in response to various cytokines<sup>14</sup>. It regulates genes that are involved in antimicrobial defense and cellular differentiation<sup>15</sup>. IL-22 can restrict the growth of *M. tuberculosis* in macrophages by enhancing phagolysosomal fusion<sup>10</sup>. However, the associations between genetic variations in the IL-22 gene and its promoter region and TB susceptibility have never been reported. In this study, we carried out this association study between the IL-22 gene, and found several SNPs in the IL-22 promoter region that influence TB susceptibility.

Our approach of SNP selection is different from the traditional way of tagging SNPs, where SNPs were usually selected based only on the linkage disequilibrium in a genomic region. The SNP selection process focuses more on efficiency, which maximizes the coverage but minimizes the cost (the number of SNPs for genotyping)<sup>16</sup>. After genotyping, one can search for the functional SNPs with close LD to the tagged SNP that shows significant associations, as we did previously in breast cancer scanning<sup>17</sup>. However, there are disadvantages of this approach. (1) The LD is usually inferred from a limited number of previously genotyped data, such as HapMap. It might be different in the current studied population. (2) It ignores the functional role of the SNP, and many tagged SNPs are in the gene desert regions, and have little biological meanings<sup>18,19</sup>. Our SNP selection approach focuses on the functional role of the SNPs, particularly in the promoter region of the gene that might affect the transcription factor binding sites. We found this simple approach very effective in identifying functional SNPs for our IL22 promoter study. Among the



**Figure 2 |** Pattern of linkage disequilibrium across the IL22 promoter region in the pooled Chinese population. The LD pattern is represented by pairwise  $D'$  values between SNPs based on genotypes from the pooled study of 479 cases and 358 controls.  $D'$  values ( $\times 100$ ) for each comparison are given in the squares. Empty squares represent  $D'$  values equal to 1.



Table 3 | Association of TB susceptibility and IL22 haplotypes and in the pooled study in Chinese (479 cases and 358 controls)

| Haplotype id | Code* | Haplotype* | Frequency | Frequencies |          | OR   | 95% CI |      | P-value <sup>‡</sup> |
|--------------|-------|------------|-----------|-------------|----------|------|--------|------|----------------------|
|              |       |            |           | Cases       | Controls |      | LCL    | UCL  |                      |
| H1           | mmmmm | CTCAA      | 0.09      | 0.08        | 0.11     | 0.75 | 0.52   | 1.07 | 0.11                 |
| H2           | mmMmm | CTTAA      | 0.01      | 0.004       | 0.03     | 0.04 | 0.01   | 0.35 | 2.12E-06             |
| H3           | mmmMm | CTCGA      | 0.07      | 0.07        | 0.07     | 1.19 | 0.70   | 2.02 | 0.68                 |
| H4           | mmMMm | CTTGA      | 0.32      | 0.31        | 0.32     | 1.32 | 0.92   | 1.90 | 0.69                 |
| H5           | MMMMM | TATGG      | 0.49      | 0.52        | 0.46     | 1.49 | 1.05   | 2.13 | 0.01                 |

\*M refers to major allele, and m refers to minor allele. \*SNPs are in the order of [rs2227485, rs2227483, rs2227478, rs2227473, and rs2227472]. <sup>‡</sup>current haplotype against the rest.

thirteen regulatory SNPs we selected, five (38.5%) showed significant association with TB susceptibility. This number is very high compared with non-regulatory SNPs, of which we selected seven SNPs, with none of them showing significant association. However, the effectiveness of this approach on other genes is unknown and yet to be tested.

The SNPs we found are biologically meaningful. The SNP rs2227473 alters the putative binding sites of several transcription factors from multiple databases, including Cgd2\_3490 and PF14\_0633, from the UniPROBE database, and RCGCANGCGY, and TCCCRNNRTG from the Jaspas database. Both Cgd2\_3490 and PF14\_0633 are hypothetical transcription factors containing AP2 domain. Cgd2\_3490 is from *Cryptosporidium parvum*<sup>20,21</sup> and PF14\_0633 is from *Plasmodium falciparum*<sup>22,23,24</sup>. We propose that the two proteins from pathogens can target the host genes and regulate IL-22's expression. The variations in the binding sites of the two proteins provide a differential way of regulation. The two motifs, RCGCANGCGY and TCCCRNNRTG, were discovered by scanning the promoters of all human genes for conserved motifs<sup>25</sup>. RCGCANGCGY matches the binding consensus of transcription factor NRF1 (nuclear respiratory factor 1), which binds to the cytochrome c promoter<sup>26</sup>, and is essential for cell survival in oxidative stress inducing agents<sup>27</sup>.

Consistent with previous research that patients with tuberculosis have significantly lower Th22 response than healthy donors<sup>28</sup>, our functional assay indicated that PBMCs from individuals carrying rs2227473G allele in IL-22, which is associated with susceptibility to tuberculosis, produce significantly lower IL-22 in response to polyclonal and Mtb antigen stimulation. Although we are currently uncertain with the mechanism of predisposition to tuberculosis in individuals with the rs2227473G allele, it probably lies at the level of the macrophage as well as lung epithelial cell. The macrophage is not only the most important effector of immunity to kill infected Mycobacterium tuberculosis, but also the niche to host them. It has been well recognized that Mycobacterium tuberculosis resides in phagosomes within macrophages, and killing these mycobacteria tuberculosis requires fusion of the phagosome with a lysosome.

Importantly, Dhiman et al demonstrated that human IL-22 inhibits growth of mycobacterium tuberculosis by enhancing phagolysosomal fusion<sup>10</sup>. While there is no evidence to show the effect of IL-22 on epithelial cells to kill Mycobacterium tuberculosis, IL-22 has been known to enhance lung epithelial cells to produce  $\beta$ -defensin<sup>15,29</sup>, which is beneficial to clear infected Mycobacterium<sup>30</sup>. Besides, several studies have shown that IL-22 participates in the protective immunity against tuberculosis<sup>28,31,32</sup>. Taken together, our findings indicate that rs2227473 polymorphism controls IL-22 response upon Mycobacterium tuberculosis infection, which in turn determines, at least in part, the outcome of infection.

In summary, our study demonstrates an association between polymorphisms in the promoter of IL-22 gene and TB susceptibility. The rs2227473 polymorphism and associated SNPs in the promoter region of the IL-22 gene are associated with decreased susceptibility to pulmonary TB in Chinese. The 'TGTAG' haplotype, which presents in 45.6% of the Chinese population, is associated with increased susceptibility to TB. The rs2227473 polymorphism affects IL-22's protein expression. Our SNP selection strategy focusing on regulatory SNPs is effective in identifying susceptible loci for the IL22 gene.

## Methods

**Subjects and samples.** Patients with different manifestations of active TB were recruited at clinics in the Shenzhen Third People's Hospital and Shenzhen Polytechnic College in Shenzhen, China. Healthy adults with no history of TB disease were also recruited as the control group. All participants had received bacillus Calmette-Guérin (BCG) vaccination at birth. We used M. tuberculosis-specific IFN- $\gamma$  enzyme-linked immunospot (ELISPOT) assay to exclude LTBI from healthy donors, and only those who were ELISPOT negative were selected as our healthy controls. The diagnosis of tuberculosis was based on mycobacterium tuberculosis examination, clinical symptoms, and chest X-ray examination as described before<sup>13</sup>. Samples collected before Feb., 2010 were used as our experimental cohort (479 cases and 358 controls) and samples collected after that date were used as our validation cohort (413 cases and 241 controls). The statistics of age distributions and male/female ratios are listed in **Supplementary Table S3**. The study obtained ethical approval from the Institutional Review Board of the Shenzhen Third People's Hospital, and informed written consent was obtained from all the patients. Clinical specimens from patients with TB were collected within one week after anti-TB treatment. Whole blood was collected by venipuncture from the populations mentioned above.

Table 4 | TB susceptibility and polymorphism of IL-22 in the pooled data analysis of association studies for different age groups of Chinese

| Age  | Sample counts |          | SNP id    | MAF   |          |    | Minor allele | OR    | 95% CI |         | P-value |
|------|---------------|----------|-----------|-------|----------|----|--------------|-------|--------|---------|---------|
|      | Cases         | Controls |           | Cases | Controls | OR |              |       | LCL    | UCL     |         |
| ≤ 25 | 147           | 95       | rs2227472 | 0.386 | 0.572    | A  | 0.469        | 0.313 | 0.703  | 2.2E-04 |         |
|      |               |          | rs2227473 | 0.120 | 0.169    | A  | 0.669        | 0.376 | 1.192  | 0.171   |         |
|      |               |          | rs2227478 | 0.166 | 0.209    | C  | 0.751        | 0.466 | 1.211  | 0.240   |         |
|      |               |          | rs2227483 | 0.393 | 0.566    | T  | 0.496        | 0.313 | 0.785  | 2.6E-03 |         |
|      |               |          | rs2227485 | 0.389 | 0.586    | C  | 0.448        | 0.303 | 0.663  | 5.2E-05 |         |
| > 25 | 332           | 263      | rs2227472 | 0.502 | 0.496    | G  | 1.025        | 0.802 | 1.309  | 0.8459  |         |
|      |               |          | rs2227473 | 0.088 | 0.134    | A  | 0.624        | 0.417 | 0.624  | 0.0208  |         |
|      |               |          | rs2227478 | 0.141 | 0.185    | C  | 0.724        | 0.528 | 0.991  | 0.0435  |         |
|      |               |          | rs2227483 | 0.486 | 0.515    | T  | 0.892        | 0.680 | 1.169  | 0.4069  |         |
|      |               |          | rs2227485 | 0.500 | 0.492    | C  | 1.034        | 0.813 | 1.315  | 0.7877  |         |



**SNP selection.** Since we focus on the SNPs with regulatory roles, we are particularly interested in two types of SNPs: (a) the SNP that is in a putative transcription factor binding site, and its two alleles that alter the binding scores, and (b) the SNP that is in a putative microRNA target site and the two alleles that alter the miRNA-target interaction scores.

To search for the SNPs with alleles that alter the putative transcription factor binding sites, we scanned the promoter region of the gene with annotated Position Weight Matrices (PWMs) in Jaspas<sup>33</sup>, UniPROBE<sup>34</sup> and TRANSFAC<sup>35</sup> databases. We first obtained the gene coordinates (txStart and txEnd) for human genome hg18 (NCBI36) from the RefSeq table of UCSC table browser (<http://genome.ucsc.edu/>). For the annotated SNPs in dbSNP129 that are within 2000 bp upstream to 500 bp downstream of the IL-22 gene, we extracted their flanking sequence (+-25bp) from the dbSNP website (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). For each SNP, we kept the flanking sequence the same and changed the alleles, thus obtaining one sequence for each allele and forming a sequence set. We then used a PWM\_SCAN algorithm<sup>36</sup> to scan each sequence in the set to test whether it had a putative binding site (PBS), with the method we described in<sup>37</sup>. We considered sites with a probability score of  $p\text{-value} \leq 0.001$  as PBS. If the SNP was within the PBS and any of the two alleles had differential  $p$ -values, we calculated the ratio ( $S_r$ ) of  $p$ -values by dividing the bigger  $p$ -value by the smaller  $p$ -value. The  $S_r$  measures how the two alleles in the PBS change the binding scores between the PBS and putative binding protein. We then converted this  $S_r$  into a  $p$ -value based on a background distribution from a permutation test, based on the FastPval program<sup>38</sup>. If an SNP has  $S_r$  with  $p\text{-value} < 0.01$ , we considered this SNP as functional and thus selected it for genotyping.

To search for the SNPs with alleles that alter the putative microRNA target sites, we first searched the PITA database for the microRNAs that could target the 3' UTR region of the IL-22 gene. For the SNP that is in the target site, we obtained its flanking sequence (+-50bp) from the UCSC genome browser according to its genomic location. We then changed the polymorphism site and obtained one sequence for each of the alleles. First, we tested whether the change of alleles also changed the microRNA-target interaction by RNAhybrid. The RNAhybrid calculates the minimum hybridization energy between the microRNA and its target. If the alleles in the polymorphism site change this energy significantly, it might enhance or repress the regulatory function of the RNA, and thus plays a role in the disease. Second, we used RNAfold to calculate the thermodynamics of the putative target site. If the alleles change the thermodynamics of the target, this might affect the chance of the microRNA binding to its target, and thus affect the function of the microRNA. SNPs with statistical significance in either target thermodynamics or microRNA-target interaction were selected for genotyping.

In addition to the regulatory SNPs, we also used the traditional tagging procedure to select SNPs that cover the IL-22 gene<sup>39</sup>. We chose the CHB+JPT HapMap panel, included our regulatory SNPs, chose  $r^2 = 0.8$  and used default settings for all other parameters. We also selected SNPs in the IL-22 gene region that are reported to be associated with disease. The functional, tagged, and selected SNPs are shown in **Supplementary Table S4**. The consensus of the putative binding sites of each selected SNP are shown in **Supplementary Table S2**.

**SNP genotyping.** Genomic DNA was extracted with a DNA isolation kit (Qiagen Inc., Germany), following the manufacturer's instruction. SNPs were typed by a high-throughput Sequenom® genotyping platform (San Diego, CA, USA). The genotypes were determined by a Homogenous Mass EXTEND assay. The Mass ARRAY AssayDesign software was used to design allele-specific extension primers. The primers designed for each selected SNP are listed in **Supplementary Table S5**. The genotyping for the experimental cohort was carried out in Shenzhen, and genotyping for the validation cohort was carried out in both Shenzhen and Hong Kong, with concordance >98%.

**Statistical analysis.** We used PLINK v1.07 and Haploview v4.0 for all the analysis undertaken. At each stage, SNPs that failed the Hardy-Weinberg equilibrium ( $p\text{-value} < 0.05$ ) were removed from further analysis. We used allelic association tests to compare the frequency of SNP alleles in cases and controls. OR and 95% CI were estimated using logistic regression models. We performed allelic association tests for different age groups ( $\leq 25$  and  $> 25$ ). The haplotype association test was performed in the region surrounding the rs2227473 marker. SNPs or haplotypes with  $p\text{-values} < 0.05$  were considered to be significant.

**Quantifications of IL-22 under specific and non-specific stimuli of PBMCs.** 5ml peripheral blood samples of TB patients were collected and PBMCs were isolated by the Ficoll lymphocyte separation medium. PBMCs were then spread into a 96-well plate at the cell density of  $4 \times 10^5$ , and 1  $\mu\text{g/ml}$  of monoclonal antibodies anti-CD3 and anti-CD28 were added for non-specific stimulation, and 20  $\mu\text{g/ml}$  of Mycobacterium tuberculosis (Mtb) lysate were added for specific stimulation. The stimulated PBMCs were incubated at 37°C, 5% CO<sub>2</sub> for 96h. Cell supernatants were used for IL-22 protein quantification with ELISA.

- The Henry J. Kaiser Family Foundation, T.H.J.K.F., 2010 (<http://www.kff.org/globalhealth/upload/7883-02.pdf>).
- Goldsack, L. & Kirman, J. R. Half-truths and selective memory: Interferon gamma, CD4(+) T cells and protective memory against tuberculosis. *Tuberculosis (Edinb)* **87**, 465–473 (2007).
- Yim, J. J. & Selvaraj, P. Genetic susceptibility in tuberculosis. *Respirology* **15**, 241–256 (2010).
- Lio, D. *et al.* Genotype frequencies of the +874T->A single nucleotide polymorphism in the first intron of the interferon-gamma gene in a sample of Sicilian patients affected by tuberculosis. *Eur J Immunogenet* **29**, 371–374 (2002).
- Rossouw, M., Nel, H. J., Cooke, G. S., van Helden, P. D. & Hoal, E. G. Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. *Lancet* **361**, 1871–1872 (2003).
- Tso, H. W., Lau, Y. L., Tam, C. M., Wong, H. S. & Chiang, A. K. Associations between IL12B polymorphisms and tuberculosis in the Hong Kong Chinese population. *J Infect Dis* **190**, 913–919 (2004).
- Thye, T. *et al.* Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet* **42**, 739–741 (2010).
- Khader, S. A. *et al.* IL-23 and IL-17 in the establishment of protective pulmonary CD4+ T cell responses after vaccination and during Mycobacterium tuberculosis challenge. *Nat Immunol* **8**, 369–377 (2007).
- Scriba, T. J. *et al.* Distinct, specific IL-17- and IL-22-producing CD4+ T cell subsets contribute to the human anti-mycobacterial immune response. *J Immunol* **180**, 1962–1970 (2008).
- Dhiman, R. *et al.* IL-22 produced by human NK cells inhibits growth of Mycobacterium tuberculosis by enhancing phagolysosomal fusion. *J Immunol* **183**, 6639–6645 (2009).
- Aujla, S. J. *et al.* IL-22 mediates mucosal host defense against Gram-negative bacterial pneumonia. *Nat Med* **14**, 275–281 (2008).
- Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nat Med* **14**, 282–289 (2008).
- Chen, X. *et al.* Reduced Th17 response in patients with tuberculosis correlates with IL-6R expression on CD4+ T Cells. *Am J Respir Crit Care Med* **181**, 734–742 (2010).
- Wolk, K. *et al.* IL-22 increases the innate immunity of tissues. *Immunity* **21**, 241–254 (2004).
- Wolk, K. *et al.* IL-22 regulates the expression of genes responsible for antimicrobial defense, cellular differentiation, and mobility in keratinocytes: a potential role in psoriasis. *Eur J Immunol* **36**, 1309–1323 (2006).
- Sham, P. C. *et al.* Combining functional and linkage disequilibrium information in the selection of tag SNPs. *Bioinformatics* **23**, 129–131 (2007).
- Hunter, D. J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics* **39**, 870–874 (2007).
- Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362–9367 (2009).
- McClellan, J. & King, M. C. Response: Why It Is Time to Sequence. *Cell* **142**, 353–355 (2010).
- Heiges, M. *et al.* CryptoDB: a Cryptosporidium bioinformatics resource update. *Nucleic Acids Res* **34**, D419–D422 (2006).
- KEGG, 2010 (<http://www.genome.jp/kegg/>).
- Aurrecochea, C. *et al.* PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* **37**, D539–543 (2009).
- De Silva, E. K. *et al.* Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc Natl Acad Sci U S A* **105**, 8393–8398 (2008).
- PlasmoDB, 2010 (<http://plasmodb.org/plasmo/>).
- Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338–345 (2005).
- Evans, M. J. & Scarpulla, R. C. NRF-1: a trans-activator of nuclear-encoded respiratory genes in animal cells. *Genes Dev* **4**, 1023–1034 (1990).
- Kwong, M., Kan, Y. W. & Chan, J. Y. The CNC basic leucine zipper factor, Nrf1, is essential for cell survival in response to oxidative stress-inducing agents. Role for Nrf1 in gamma-gcs(l) and gss expression in mouse fibroblasts. *J Biol Chem* **274**, 37491–37498 (1999).
- Scriba, T. J. K. B., Abrahams, D. A., Isaacs, F., Hofmeister, J., Black, G., Hassan, H. Y., Wilkinson, R. J., Walz, G., Gelderbloem, S. J., Mahomed, H., Hussey, G. D., Hanekom, W. A. Distinct, specific IL-17- and IL-22-producing CD4+ T cell subsets contribute to the human anti-mycobacterial immune response. *J Immunol* **180**, 1962–1970 (2008).
- Whittington, H. A., Armstrong, L., Uppington, K. M. & Millar, A. B. Interleukin-22: a potential immunomodulatory molecule in the lung. *Am J Respir Cell Mol Biol* **31**, 220–226 (2004).
- Rivas-Santiago, B., Contreras, J. C., Sada, E. & Hernandez-Pando, R. The potential role of lung epithelial cells and beta-defensins in experimental latent tuberculosis. *Scand J Immunol* **67**, 448–452 (2008).
- Yao, S. *et al.* Differentiation, distribution and gammadelta T cell-driven regulation of IL-22-producing T cells in tuberculosis. *PLoS Pathog* **6**, e1000789 (2010).
- Qiu, L. *et al.* Severe tuberculosis induces unbalanced up-regulation of gene networks and overexpression of IL-22, MIP-1alpha, CCL27, IP-10, CCR4, CCR5, CXCR3, PD1, PDL2, IL-3, IFN-beta, TIM1, and TLR2 but low antigen-specific cellular responses. *J Infect Dis* **198**, 1514–1519 (2008).
- Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91–94 (2004).



34. Newburger, D. E. & Bulyk, M. L. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **37**, D77–82 (2009).
35. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* **31**, 374–378 (2003).
36. Hannenhalli, S. *et al.* Transcriptional genomics associates FOX transcription factors with human heart failure. *Circulation* **114**, 1269–1276 (2006).
37. Wang, J. W., Zhang, S. L., Schultz, R. M. & Tseng, H. Search for basonuclin target genes. *Biochemical and Biophysical Research Communications* **348**, 1261–1271 (2006).
38. Li, M. J., Sham, P. C. & Wang, J. W. FastPval: a fast and memory efficient program to calculate very low P-values from empirical distribution. *Bioinformatics* **26**, 2897–2899 (2010).
39. de Bakker, P. I. *et al.* Efficiency and power in genetic association studies. *Nat Genet* **37**, 1217–1223 (2005).

## Acknowledgments

We thank Yan Zhang, Wanling Yang and Yu Lung Lau of the University of Hong Kong for assistance in genotyping samples and critical comments on the manuscript. Financial support was provided by Eleven-Fifth Mega-Scientific Project on “prevention and

treatment of AIDS, viral hepatitis and other infectious diseases” (2008ZX10003-005,2008ZX10003-012) and National Science Foundation (30872258) of China, and Research Grants Council (778609M, N\_HKU752/10) and Research Fund for the Control of Infectious Diseases (10091262) of Hong Kong.

## Author contributions

Study design (XC, BPZ, JW); data acquisitions and genotyping (GZ, YZ, MZ, BHZ, LW, XZ, JZ); data analysis (LC, GZ, XC, JW); article writing (XC, JW).

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

**How to cite this article:** Zhang, G. *et al.* An SNP selection strategy identified IL-22 associating with susceptibility to tuberculosis in Chinese. *Sci. Rep.* **1**, 20; DOI:10.1038/srep00020 (2011).