



OPEN

SUBJECT AREAS:

ENVIRONMENTAL
SCIENCES

RISK FACTORS

Received
6 June 2014Accepted
12 September 2014Published
9 October 2014Correspondence and
requests for materials
should be addressed to
S.T. (taos@pku.edu.cn)

Global lung cancer risk from PAH exposure highly depends on emission sources and individual susceptibility

Huizhong Shen, Shu Tao, Junfeng Liu, Ye Huang, Han Chen, Wei Li, Yanyan Zhang, Yuanchen Chen, Shu Su, Nan Lin, Yinyin Xu, Bengang Li, Xilong Wang & Wenxin Liu

Laboratory for Earth Surface Processes, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China.

The health impacts of polycyclic aromatic hydrocarbons (PAHs), the most concerning organic pollutants, depend not only on the locations and strengths of emission sources, but also on individual susceptibility. Moreover, trans-boundary transport makes them a global concern. In this study, a comprehensive analysis of the global health impacts of polycyclic aromatic hydrocarbons (PAHs) in ambient air is presented. Model resolution is critical in exposure modelling. Globally, incremental lifetime lung cancer risk (*ILCR*) induced by ambient PAH exposure is 3.1×10^{-5} . If the individual susceptibility was not taken into consideration, the overall risk would be underestimated by 55% and the proportion of highly vulnerable population would be underestimated by more than 90%. Emphasizing on individual susceptibility, our study provides an instrumental revision of current risk assessment methodology. In terms of lung cancer risk, the most important sources are combustion of biomass fuels (40%) and fossil fuels (14%) in the residential/commercial sector, coke (13%) and aluminium (12%) production, and motor vehicles (9%). PAHs can travel long distance globally especially within the Eurasian continent. Still, the risk is dominantly contributed by local.

It is well known that exposure to polycyclic aromatic hydrocarbons (PAHs) through various routes is associated with cancer^{1–3}. Increased lung cancer risks from both occupational and environmental exposure to PAHs were observed and well documented by previous studies^{2,4,5}. Consequently, benzo[a]pyrene (BaP), one of the high molecular weight PAHs, has been classified into the group of most carcinogenic agents by the International Agency for Research on Cancer⁶. One of the most important pathways by which humans are exposed to PAHs is inhalation, and the majority of PAHs present in the atmosphere come from incomplete combustion or from pyrolysis^{4,7}. Different PAH emission sources contribute to human exposure to PAHs to different extents because of their different emission strengths and proximities to people⁸. The quantification of source–receptor relationships is essential for assessing the health risks associated with different sources and for developing cost-effective abatement strategies.

Their widespread emissions and potential for long-range transport lead to the fact that PAHs can be detected all around the world, including in polar regions⁹. Using global chemical transport models, the fates of PAHs have been quantitatively characterized^{10–12}. However, the spatial resolutions of these models are usually as large as 1–5° longitude and latitude limited by computing load and meteorological information, which means that models can only provide averaged values within grids of ten thousand to several hundred thousand square kilometers. Concentrations calculated at such coarse resolutions are not suitable for risk assessments because population densities vary at a much finer spatial resolution. This scale mismatch was also found to be problematic in a previous effort to model population exposure to black carbon, and a simple downscaling method was used to generate highly resolved gridded concentrations¹³.

Evidence from molecular epidemiologic studies has indicated that individual susceptibility plays an important role in cancer development in humans under environmental stress¹⁴. However, the influences of susceptibility on cancer risks in populations grouped by gender, age, and genetic heritage are difficult to characterize quantitatively. It is a challenge to distinguish the relative contributions of the exposure dose and individual susceptibility to the risks associated with PAHs¹⁴. Based on the result of an epidemical study conducted in Xuanwei, China¹⁵, and relationship between DNA adduct level and number of at-risk alleles, an attempt has been made in a study in which the risks of PAH exposure inducing lung cancer were assessed with consideration of susceptibility associated with individual genetic polymorphisms in the Chinese population¹⁶.

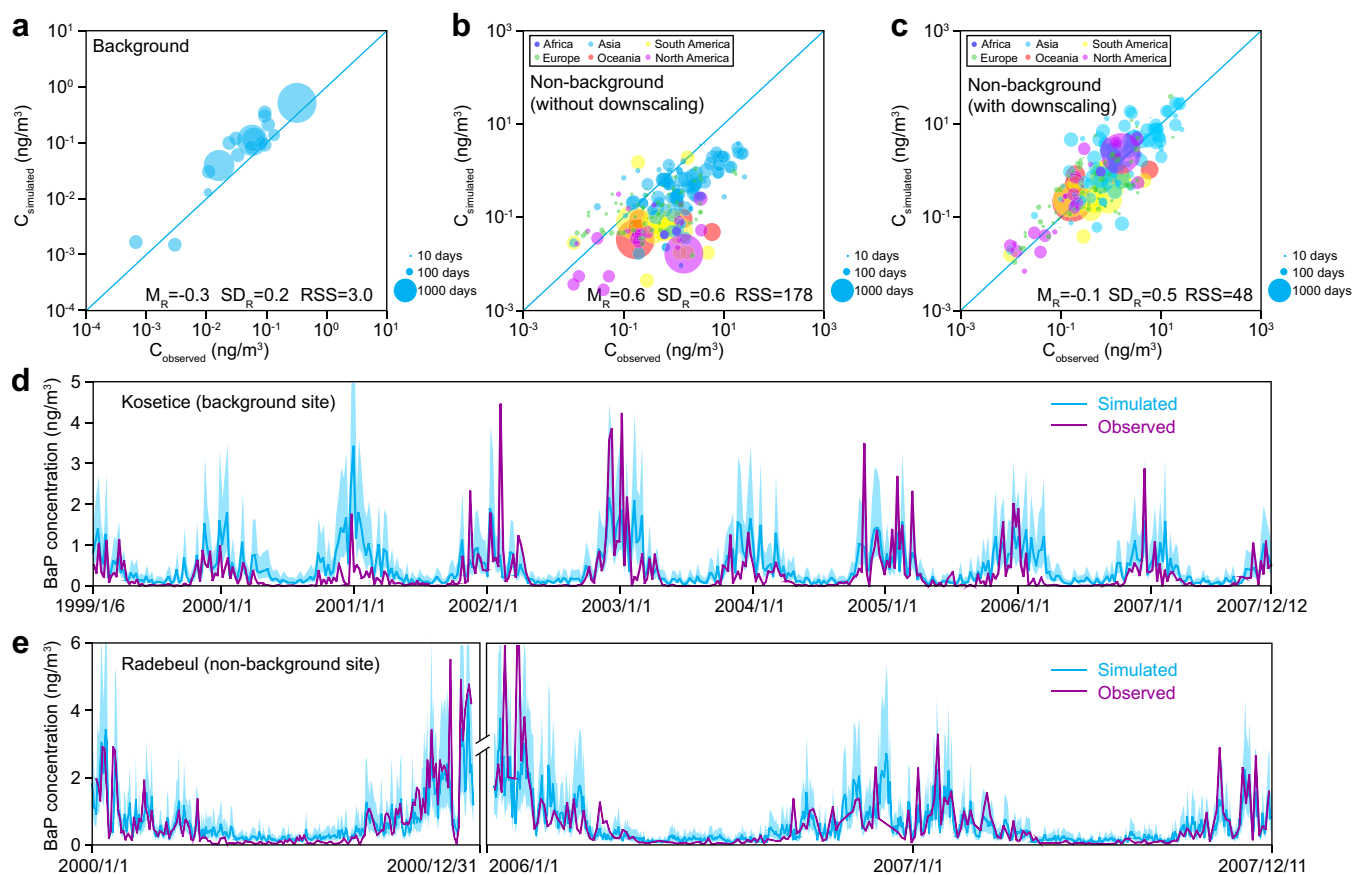


Figure 1 | Validation of the model calculated near-surface BaP concentrations against observations in the field. The downscaling is conducted for non-background grids with strong spatial variations. (a) Validations at 18 background sites. (b and c) Validations at 236 non-background sites without and with downscaling, respectively. Sites in different regions are marked with different colours. The areas of the bubbles are proportional to observation durations, so more reliable observations are carried in larger bubbles. 1 : 1 lines are drawn in blue. Residuals are defined as the differences between log-transformed observations and calculations, and the calculated means (M), standard deviations (SD), and square sums (RSS) of residuals are listed in (a), (b) and (c). (d and e) Time series at Kosetice background and Radebeul non-background stations. The calculated and observed concentrations are shown as blue and purple lines, respectively. The model calculations at Radebeul non-background site are downscaled. The model uncertainties as 95% confidence intervals are shown by shaded areas. The comparisons of time series for all other six background stations and one non-background station can be found in the Supplementary Figs. S3–S5.

Recently, a $0.1^\circ \times 0.1^\circ$ global PAH emission inventory, with 69 detailed source types, has been developed⁷. Using this product, we developed a global chemical transport model for BaP (as a indicator for PAH risk assessment) and assessed the risks in global population of lung cancer being induced by the inhalation of PAHs. We used a downscaled high-resolution BaP concentration map, distinguishing the contributions of different emission sources to the risks, and quantified the long-range transport of BaP across continents. We also determined the variation of individual susceptibility and quantified the influence of individual susceptibility on overall risk against the influence of exposure dose on a global scale.

Results

Global Near-Surface BaP Concentrations and Population Exposure. We developed a BaP module incorporated in MOZART-4¹⁷ to simulate the global transport of BaP at a resolution of $1.875^\circ \times 1.895^\circ$ with 28 vertical layers, from 1997 to 2007 (see the Methods). Using the three quartiles of the emissions at all grids, we generated the best estimates and uncertainty ranges (R_{50} , defined as the interquartile range) of global BaP concentrations. The model predictions agreed well with observations at background sites, but underestimated considerably at non-background sites (Fig. 1a,b). This kind of underestimation was familiar in previous studies^{12,13,16} and was likely caused by scale mismatch, because most non-background

sites were located in cities, while the model provided average concentrations of grid cells, covering both cities and less-contaminated surrounding areas. Two examples of such mismatch are shown in Supplementary Fig. S2. To rectify the underestimation, we downscaled the calculated concentrations to $0.1^\circ \times 0.1^\circ$ based on emission density and wind direction/frequency/speed. With this new method, we found a good agreement between the calculations and observations (Fig. 1c). In addition, modelled time series at several stations were satisfactorily validated (Fig. 1 d,e and Supplementary Figs. S3–S5).

The near-surface BaP concentrations are mapped in Fig. 2a. Grid BaP concentrations over land vary extensively with the three quartiles of 0.00032, 0.0055, and 0.046 ng/m^3 , respectively. East Asia and South Asia are the major hot regions. Monthly maps suggest a strong seasonality with higher concentrations in winter due to extra emissions from heating and slow degradation (Supplementary Fig. S6). The temperature dependence of BaP persistency is confirmed by the strong latitudinal and seasonal variations of BaP's air loss rate (Supplementary Fig. S6). Because of the spatial co-variation between emission and population, the geographical distribution of population-weighted concentrations as risk indicators (Fig. 2b) shows intensified risk in the hot regions. The three quartiles of the population exposure concentrations are 0.18, 0.57 and 1.65 ng/m^3 , respectively. The median population exposure concentrations in

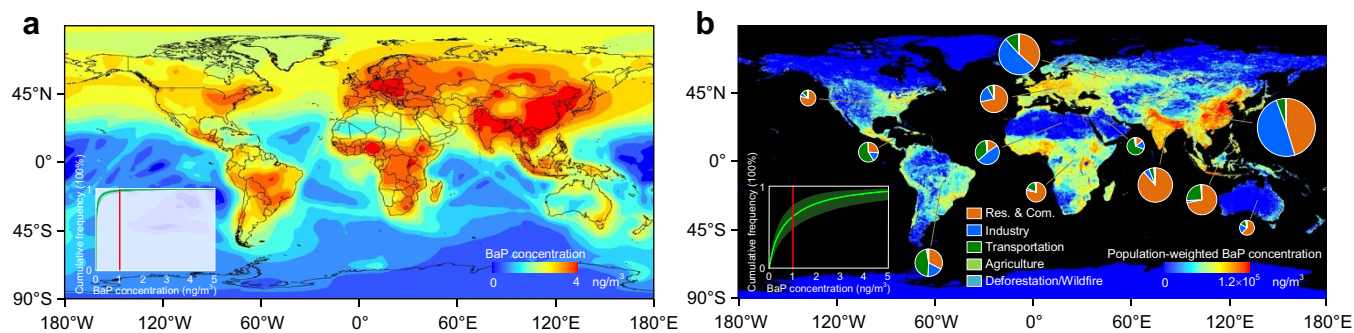


Figure 2 | Geographical distributions of BaP concentrations in 2007. (a) Ambient air BaP concentrations calculated from MOZART-4. (b) Population-weighted and spatially-downscaled BaP concentrations as indicators of population exposure risk. The relative contributions of various source sectors are shown as pie charts for individual regions. The areas of the pie charts are proportional to the regional mean exposure concentrations. The 12 regions are defined by the Task Force on Hemispheric Transport of Air Pollution¹⁸ (Supplementary Table S10); the Arctic, Antarctic, and non-Arctic/Antarctic oceans are not shown because of very low values. Cumulative distributions of grid concentrations are also shown together with the European Union target value of 1 ng/m^3 in these two maps, respectively. Uncertainties at 95% confidence intervals are shown by shaded areas. This figure was generated using ESRI® ArcMap, Microsoft® Office Excel, and Microsoft® Office Word^{19,20}.

East and South Asia are as high as 2.01 and 0.69 ng/m^3 , respectively. Emissions from residential and industrial sectors dominated the exposure because of high emission factors and proximity to people (Fig. 2b).

Cumulative frequency distributions of BaP concentrations are also shown in Fig. 2 (globe) and Supplementary Table S1 (individual regions). Globally, there is 2.2% (1.8–2.7% as R_{50} hereafter) of land area with ambient air concentrations exceeding the EU target value (1 ng/m^3)²¹. However, 36.9% (32.0–41.7%) of the world's population live in these areas. Among them, 80.3% live in East, South, and Southeast Asia, where 72.2, 40.1, and 35.4% of the population, respectively, are exposed to annual mean ambient BaP concentrations above 1 ng/m^3 . Even with a relatively lenient national standard²² of 2.5 ng/m^3 , the rate of excess in China is as high as 34.1%.

Risk Assessment and Influence of Individual Susceptibility. We used BaP as an indicator for PAH risk assessment. To be precautionary, the 95% upper bound of the slope of a dose-response regression curve is commonly adopted as the cancer slope^{23,24}. In this study, however, we used the maximum likelihood, instead of the upper bound, to derive an unbiased best estimate of *ILCR* and Population Attributable Fraction (PAF). Meanwhile, uncertainty due to variation in the dose-response relationship is explicitly quantified in a probabilistic model. In addition, genetic susceptibility is characterized based on frequency distributions generated from the lung cancer-related genetic polymorphism data for the major human races (see the Methods).

Globally, overall *ILCR* and PAF for lung cancer risk induced by inhalation exposure to ambient PAHs are 3.1×10^{-5} (1.6×10^{-5} – 5.9×10^{-5}) and 0.13% (0.07–0.25%), respectively, taking individual susceptibility into consideration (see the Methods). The R_{50} of the log-scaled variation ($R_{50, \log}$) in individual susceptibility (0.81) is close to that of the exposure dose (0.96), suggesting that the relative contribution of individual susceptibility to the overall variation of the risk is comparable to that of the exposure at global scale. It is expected that on regional or local scale, difference in individual susceptibility can play a more important role in terms of risk variability than variations in exposure dose can.

Although the PAF is much lower than that associated with smoking (71%)²⁵, we found a significantly positive correlation ($p < 0.1$) between the country-specific *ILCRs* and lung cancer incidences²⁶ (Supplementary Fig. S9). Normalized by smoking prevalence²⁷, the correlation was more significant ($p < 0.01$, Supplementary Fig. S9). In the smoking effect normalized regression model, 12% of the variation (R^2) in lung cancer incidences can be explained by the variation in the calculated *ILCR* from ambient PAH exposure. Such a relation-

ship between lung cancer incidences and the calculated *ILCR* is more robust in developed countries, where 38% of the variation in lung cancer incidences is accountable by the variation in *ILCR*, likely due to well-established cancer registry systems. In addition, the geographic similarity of *ILCRs* and lung cancer mortality can be further illustrated at a sub-national scale in China (Supplementary Fig. S9). Previous study noted that lung cancer mortality rates of non-smokers in China varied widely in different regions²⁸. The spatial similarity between *ILCRs* and lung cancer mortality in this study suggests air pollution as an important factor effecting the geographic variability of non-smokers lung cancer mortality rates.

Figure 3 shows the log-scaled frequency distributions of global *ILCR* with (blue curve) and without (green curve) individual susceptibility included. The overall risk as *ILCR* would have been underestimated by 55% (1.4×10^{-5}) if individual susceptibility was not considered. With susceptibility, the cumulative curve was more platykurtic, and the population fractions at both high and low risk ends increased dramatically, as shown by the ratios of the two frequencies (brown curve). Most importantly, because of the variation in individual susceptibility, the fractions of the most vulnerable population with risk above 10^{-4} and 10^{-3} increased 1.5 (from 2% to 5%) and 9 times (from 0.03% to 0.3%), respectively. We then calculated Unit Risk (*UR*, m^3/ng), defined as the *ILCR* from exposure to per unit BaP concentration²⁹, for all countries (Supplementary Fig. S10). The large variation in *UR* among countries is caused by the difference in susceptibility, and suggests that use of a single *UR* in risk assessment can lead to considerable spatial bias. Accordingly, to keep *ILCRs* of 97.5% population below 10^{-5} , it is necessary to set a BaP guideline of 0.1 ng/m^3 globally. Based on the *ILCRs* and reported country incidence rates of lung cancer²⁶, Relative Risk (*RR*) of lung cancer from ambient PAHs exposure was calculated for each country. Globally, the overall *RR* is 1.0013. *RR* of the developing world (1.0017) is higher than that of the developed world (1.0003), which can be explained by the fact that populations in developing countries have much higher levels of exposure concentrations to ambient PAHs and lower levels of lung cancer incidence rates²⁶. Using the relative risks, we then calculated the Unit Relative Risk (*URR*)⁴ at a benchmark of $100 \mu\text{g/m}^3$ years of BaP exposure for each country. Also, we calculated the *URR* values for individual persons within each country, assuming that different people have different *URRs* due to their differences in susceptibilities. Globally, the interquartile range of country-level *URRs* is [1.8, 7.0] with the median value of 3.0. It was found that the population overall *URR* of a specific country is higher than the upper bound of interquartile range of individual *URRs*. For example, the overall *URR* of Chinese population is 1.56, whereas the interquartile range of individual *URRs* among Chinese population is merely [1.02, 1.25].

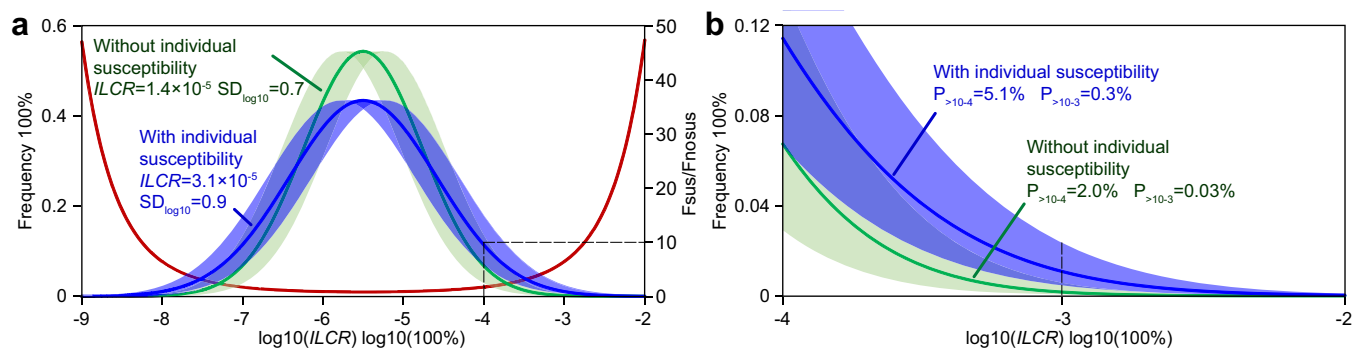


Figure 3 | Frequency distributions of global *ILCR* due to inhalation exposure to ambient PAHs. (a) The two curves show the model-calculated frequency distributions with (blue) and without (green) individual susceptibility considered. The distribution of the ratios of the two frequencies (with susceptibility to without susceptibility) is shown as the brown line, indicating accelerated increase in the probabilities at both low ($ILCR < 10^{-8}$) and high ($ILCR > 10^{-5}$) risk ends. The overall average *ILCR*s together with standard deviations, with and without susceptibility, are listed. (b) A zoomed-in version of (a) in the high risk range. The probabilities under the two distribution curves with *ILCR* higher than 10^{-4} and 10^{-3} are labelled as $P_{>10^{-4}}$ and $P_{>10^{-3}}$, respectively. Uncertainty as interquartile range is shown as the shaded area for each frequency-distribution curve.

Actually, there are only 13% of Chinese population exceeding the *URR* value of 1.56, suggesting that this small fraction of people with highest *URR*s can influence the overall *URR* of the population to a very large extent.

It was suggested that the uncertainties in quantifying dose-response relationship partly relies on a lack of quantitative information on the level of exposure³⁰. However, in this study, the uncertainty range of dose-response relationships from epidemiological surveys ($R_{50,\log} = 0.82$)³¹ is comparable to the variation of individual susceptibilities ($R_{50,\log} = 0.81$), suggesting that the uncertainties can be also raised by variation in individual susceptibilities, to which genetic polymorphisms contributes most. The adoption of individual susceptibility can not only explain the uncertainties of risk assessments but also change our insight into the risk outcomes. The air quality guideline, determined by current risk assessment methodologies, commonly with 97.5% confidence (upper bound of 95% confidence interval) in keeping population exposure risks under a specified level is actually, in our point of view, a guideline to keep exposure risks of 97.5% population under the specified level. The more susceptible groups (the rest 2.5% population) are lack of attention in policy making. Emphasizing on the high risk groups, this study provides a revised method serving as the first attempt to introduce individual susceptibility into quantitative risk assessments on a global scale. With increasing knowledge of genetic and acquired susceptible factors, this methodology is expected to be more instrumental in evaluating global burden of disease, in developing environmental and health policies, as well as in strengthening individual awareness.

Source Contributions. Several short-term modelling experiments were conducted using various source types individually to evaluate their contributions to the overall health impact. Globally, biomass fuel burning in residential/commercial sector contributes 40% of the total *ILCR*, followed by residential/commercial fossil fuel combustion (14%), coke production (13%), primary aluminium production (12%), and motor vehicles (9%) (Fig. 4a and Supplementary Table S9). The contribution of a given source to the risk depends not only on the emission strength but also on the proximity to people. Therefore, relative contributions of a given source to the total emission and to the total risk can be different. The effect of a given source is quantified by specific health effect (*SILCR*, Pg^{-1}), defined as *ILCR* caused by per unit emission. Globally, *SILCR*s for various source types vary extensively from 0.08 and 0.14 Pg^{-1} for shipping and wildfire to 13 and 17 Pg^{-1} for residential/commercial fossil fuel usage and motor vehicles (Fig. 4a and Supplementary Table S9). Motor vehicles are associated with the highest *SILCR*, due to its spatial intensification in populated areas.

*ILCR*s also vary dramatically among populations at different risk levels (Fig. 4b). For example, a small fraction (1.7%) of the population facing high risk ($ILCR > 3 \times 10^{-4}$) is largely because of exposure to emissions from coke and aluminium productions. Especially, *ILCR*s of some inhabitants near coking plants in China can reach as high as 1.5×10^{-2} . Long-term situations of these people are even worse, given their cumulative exposure to continuously high levels of ambient PAHs in the past decades, during which period beehive coking, which is among the severest PAH emission sources, had not been entirely phased out in China⁸. Emissions from residential solid fuel combustion contribute mainly to the population at risk levels between 3×10^{-8} and 1×10^{-4} . Unlike industrial sources, PAH emissions from residential sources are primarily credited to millions of residential stoves. It is socioeconomic development rather than technical progress that is critical in the abatement of emissions from this sector. The risk-specified source contributions for several representative countries are shown in Supplementary Fig. S11.

These results provide a sound scientific basis for abatement strategy formulation. Globally, residential biomass burning causes the largest overall lung cancer risk and should be the top priority in the emission abatement. On the other hand, control of emissions from motor vehicles and residential coal combustion with the highest *SILCR* is the most effective way of reducing risk. If the objective is to protect the most vulnerable people, emissions from coke and aluminium production should be addressed first. However, the specific strategies should be different among countries, depending on local emissions and risks. For example, the overall risk in Russia is dominated by primary aluminium production, while motor vehicles are responsible for the risk of vulnerable populations in Indonesia. Of course, abatement costs should also be taken into account before the action plan is formulated. It should be noted that this study assumed a lifetime exposure to outdoor PAHs. Indoor exposure was not considered due to a lack of linkage between outdoor and indoor concentrations on a global scale. Nevertheless, previous investigation revealed that compared to outdoor exposure, due to the demand for cooking and heating and poorly-ventilated conditions, indoor exposure can be much higher in some of the rural areas in developing countries, especially for women during cooking period¹⁵. In this respect, an integrated risk assessment with human activity factors will be needed considering both outdoor and indoor exposure. Such an assessment may further promote the health impacts contributed by residential/commercial sector.

Transport across continents. To characterize long-range transport of BaP among 12 regions (see the Supplementary Methods and Supplementary Table S11), short-term modelling experiments

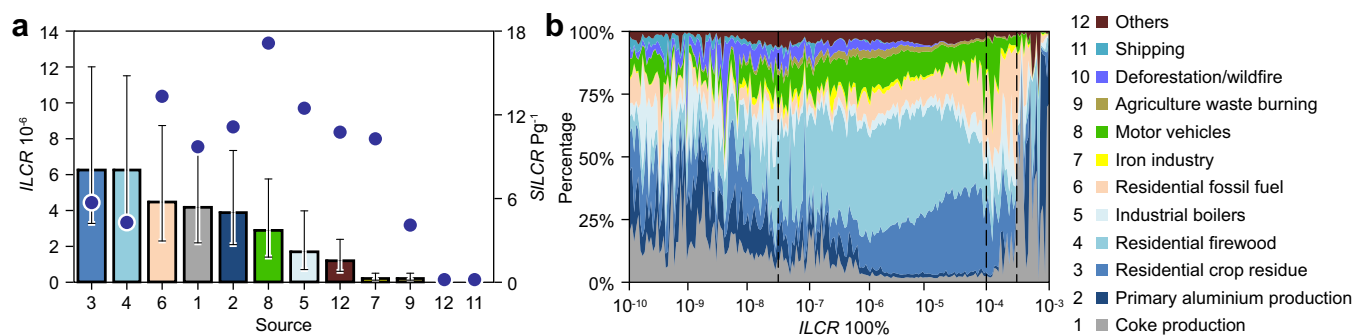


Figure 4 | Contributions of various sources to global lung cancer risks. (a) Risks caused by PAH emissions from various sources as *ILCR* (bars) and risks caused by per unit emissions of various sources as *SILCR* (dots). (b) Relative contributions of various sources to the risks for populations facing different risk levels (*ILCR*).

were conducted with emissions from a single region each time. As a result, quantities of BaP transported among regions can be quantified. The results are presented as contributions to *ILCR*. Figure 5 shows the flows of BaP including directions and quantities among regions (the red arrows) and percentages of *ILCRs* contributed by BaP from other regions (the numbers in the black pies). The sizes of the black pies are proportional to the regional *ILCRs*. It appears that interregional transport within the Eurasian continent is active, while America, Oceania, and Africa have much fewer exchanges with other continents. As a result of the westerly wind movement and lower air loss rate in high-latitude area, the transport of BaP from Western/Eastern Europe to the Former Soviet Union region represents the largest *ILCR* flow, leading to a 4.5×10^{-7} increase to local *ILCR* of the latter. The second largest movement occurs from East Asia to Southeast Asia, owing to the extraordinary emission intensity in the source region. Still, even with the active interregional transport, regional risks are predominantly caused by local emissions. The highest external contribution to local *ILCR* is merely 2.6% (Southeast Asia). The net exported risk (*NER*), defined as the difference between the exported and imported *ILCR* multiplied by regional population, is calculated for each region (Supplementary Tables S11,S12). Western/Eastern Europe ($NER = 145$), South Asia ($NER = 57$), and East Asia ($NER = 42$) are the main export regions of risk, with positive values of exported risk, while Southeast Asia ($NER = -126$), the former Soviet Union ($NER = -92$), and the Middle East ($NER = -19$) are risk passive recipients. The dynamics of the long-range transport of BaP illustrates a strong seasonality with relatively active transport in winter.

Considering rapid degradation of certain substances like PAHs, their proportions on atmospheric particles must undergo dramatic change during transport. It is estimated that mass fractions of BaP bound on atmospheric organic and black carbon decrease more than 80% (from 1.6×10^{-4} to 2.7×10^{-5}) over trans-Pacific transportation. Given the ever-changing properties, particles after long-range transport may act with very different health outcomes associated with human exposure.

Discussion

As well demonstrated in this study, model resolution is critical in exposure modelling and risk assessment. With large spatial variation of emission, coarsely resolved model will smooth detailed variation leading significant underestimation of exposure. This problem can be prevented by either using highly resolved model (e.g. 0.1°) or downscaling the model results to finer resolution as did in this study. The latter can provide an alternative to the high resolution modelling when computing load is limited. More study is needed to quantify the effects of model resolution on the results and to improve the downscaling method.

In this study, long-range transport of BaP among continents is quantified. Detailed study can be carried out to address cross-boundary transport of PAHs among countries or even sub-country administrative regions and to focus on sources of PAHs reaching vulnerable regions such as Arctic.

The sectorally-resolved emission inventory enables us to exploit the source-receptor relationship linking emissions from individual sources to population exposure risk. The outcome provides critical and quantitative information for decision makers to formulate future strategy on source abatement and risk reduction. Of course, further cost-effective analysis is also needed for the decision making process, which can not be fulfilled without the quantitative source-receptor relationship generated in this study.

One of the major findings of this study is that the lung-cancer risk would be underestimated by almost 50% if individual susceptibility is not taken into consideration in risk assessment. It is a common practice in risk assessment to use upper bound of 95% confidence interval of risks conservatively in guideline formulation. With individual susceptibility introduced in the model, the goal can be achieved more reasonably by using the statistically best estimates and generate a frequency distribution from which a given risk range can be selected. On the other hand, the susceptibility was modelled with relatively large uncertainty largely due to limited data on genetic susceptibility. In fact, the uncertainty caused by the data limitation in genetic susceptibility is comparable to those introduced by the data limitation in dose-response relationship which is constrained by limitation in epidemiological survey studies. The overall uncertainty can be reduced in the future with a full understanding of the spectrum of genetic susceptibility in the population.

Methods

Atmospheric Transport Modelling. We incorporated a PAH module into MOZART-4 (The Model for Ozone and Related Chemical Tracers, version 4)¹⁷ to simulate the global tropospheric transport of BaP, an indicator for PAHs. The modelling was performed at a horizontal resolution of 1.875° (longitude) \times 1.895° (latitude) with 28 vertical layers and 15-min temporal resolution driven by meteorology from the National Centers for Environmental Prediction (NCEP) reanalysis³².

We adopted a dual absorption scheme including black carbon (BC) and organic carbon (OC) for BaP gas/particle partitioning, and incorporated temperature-dependent partitioning coefficients^{12,33}. Gas phase reaction with hydroxyl radical (OH) was included for BaP degradation in the atmosphere¹¹, which was proved to be one of the dominant atmospheric loss processes of BaP^{34,35}. The second-order rate coefficient was set to be 1.5×10^{-10} cm³/(molec·s)¹⁰. Wet and dry depositions of particulate phase BaP followed the default settings of BC and OC aerosols in the model. Daily mean BC, OC, and OH concentrations were directly derived from MOZART-4 standard simulation. Dynamic soil/ocean-air exchanges of gaseous phase BaP, leaching and biodegradation in soils, and exchange and degradation in the ocean were incorporated after Zhang et al¹¹.

We updated a high resolution PAH emission inventory for 2007 (PKU-PAH-2007)⁸ by replacing geometric means of emission factors with arithmetic means³⁶. As a result, the global BaP emission in 2007 became 4.6 Gg. For the other years, $0.1^\circ \times 0.1^\circ$ resolution emission inventories were generated based on annual emissions of

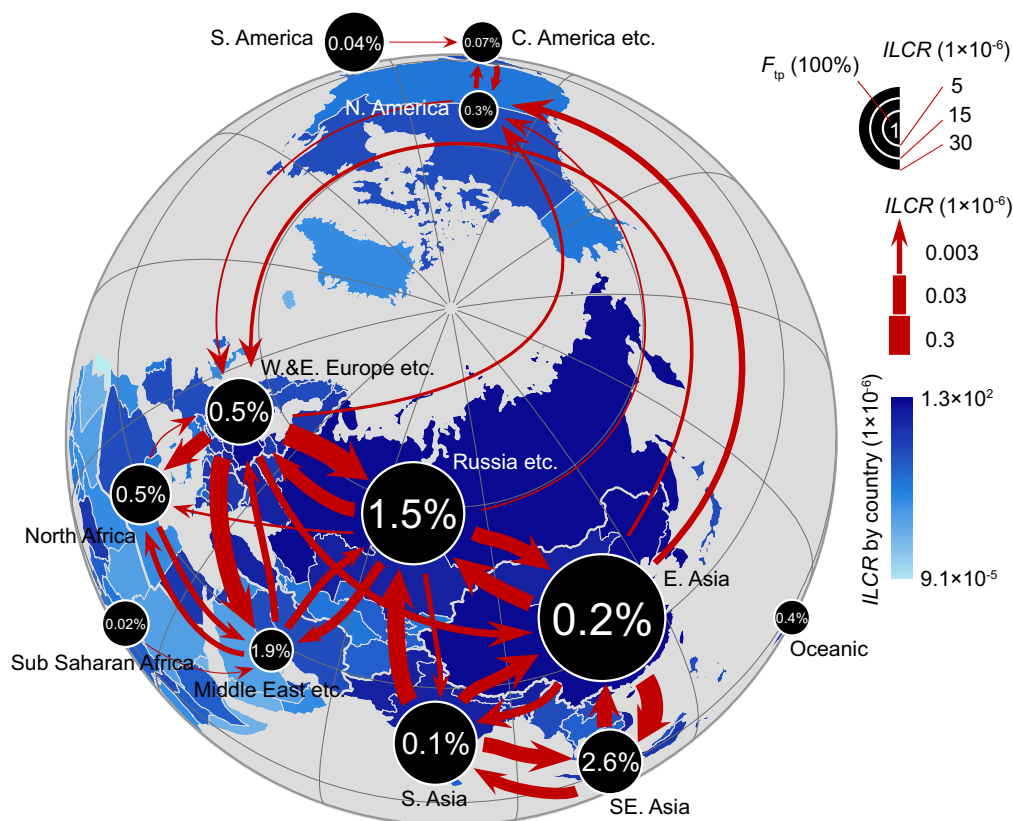


Figure 5 | Interregional transport of PAH exposure risk. The sizes of the black pies are proportional to the log-transformed overall *ILCR* of each region. The numbers in the middle (F_{ip}) are percentages of regional *ILCR* contributed by local exposure to PAHs transported from other regions. The red arrows indicate the directions of the interregional transport of BaP, and the thicknesses of the arrow shafts are proportional to the enhancement of *ILCR* due to interregional transport in log-scale. Unpopulated regions, including the Arctic, Antarctic, and non-Arctic/Antarctic oceans, and transport with *ILCR* influence less than 1×10^{-9} are omitted. The calculated *ILCRs* for individual countries are shown by colour codes in the background. This figure was generated using ESRI® ArcMap, Microsoft® Office Excel, and Microsoft® Office Word^{19,20}.

individual countries and the spatial distribution of PAH emissions in 2007, assuming that the spatial distribution patterns of the emissions did not change significantly. Monthly variation of BaP emissions from the residential sector was predicted using heating day and power-function-based heating degree day as independent variables³⁴. Monthly variation of BaP emissions from deforestation/wildfire and agricultural waste burning were from the GFED dataset³⁸. For all other sources, it was assumed that seasonal variations were not significant.

We conducted a long-term simulation for the period from 1998 to 2007, and two sets of short term modelling experiments for the year 2007 with spin-up time of one year. The short term experiments were either for 12 individual source categories globally (Supplementary Table S8) or total emissions of 12 individual regions (Supplementary Fig. S12). The world regions were defined based on the definition of the Task Force on Hemispheric Transport of Air Pollution¹⁸ (Supplementary Table S10).

Model Validation and Downscaling. The model-calculated BaP concentrations at the surface layer were validated against the field measured BaP concentrations at more than 200 stations around the world. The observation sites were classified into three categories of background, non-background and time series, which were used separately for the validation. Locations of background and non-background sites applied for model validation are demonstrated in Supplementary Fig. S1.

To rectify the bias due to the scale mismatch, we down-scaled the model-calculated concentrations from $1.875^\circ \times 1.895^\circ$ to $0.1^\circ \times 0.1^\circ$. A weighting factor (W_i) for the i th $0.1^\circ \times 0.1^\circ$ receiving grid was the sum of the contributions of the emissions from all $0.1^\circ \times 0.1^\circ$ emission grids within nine $1.875^\circ \times 1.895^\circ$ model grids (one covering the $0.1^\circ \times 0.1^\circ$ receiving grid and the eight surrounding it). It was derived based on $0.1^\circ \times 0.1^\circ$ emission densities and wind field (direction, frequency, and speed) within the space of the nine $1.875^\circ \times 1.895^\circ$ grids, using equation (1)³⁹:

$$W_i = \sum_{j=1}^n \frac{2.03Q_j f_j e^{-r_d t_{ji}}}{u_j \sigma_{zj} x_{ji}} \quad (1)$$

where Q_j (ng/s) is emission density of the j th emission grid ($0.1^\circ \times 0.1^\circ$); f_j (dimensionless) and u_j (m/s) are wind frequency (0–1) and speed at the direction from 1 to 16 (N, NNE, NE, NEE, E, SEE, SE, SSE, S, SSW, SW, NWW, NW, and NNW) in the j th emission grid, which were derived from the MOZART-4 wind field; r_d (1/s) is the BaP

degradation rate in the receiving grid derived from the MOZART-4 simulation; t_{ji} (s) and x_{ji} (m) are distance and transport time from the j th emission grid to the i th receiving grid; σ_{zj} (m) is the vertical standard deviation of the concentrations, and is determined using equation (2):

$$\sigma_z^2 = 2K_z t_L \left\{ \tau - (1 - e^{-\tau}) - \frac{1}{2}(1 - e^{-\tau})^2 \right\} \quad (2)$$

where K_z is the vertical eddy diffusivity; t_L is the Lagrangian time scale of atmospheric turbulence; τ is the travel time in units of t_L ($\tau = t/t_L$). Values of $1 \text{ m}^2/\text{s}$ and $1 \times 10^4 \text{ s}$ are used for K_z and t_L , respectively, on a global scale⁴⁰. Finally, the model-calculated concentration of each $1.875^\circ \times 1.895^\circ$ model grid was disaggregated to $0.1^\circ \times 0.1^\circ$ using W_i as proxy. It should be noted that this method leads to discontinuity into the downscaled map, since the transport mechanism and environmental behaviours are only simply performed during the downscaling processes. For air pollutants with sufficient global surface measurements and remote sensing, regression analysis to combine these data with air transport models is the primary choice to carry out downscaling processes⁴¹. For regional simulations, chemical transport models with higher resolution are preferred.

To emphasize the higher health impact of BaP concentrations in densely populated areas, a population-weighted BaP concentration of each $0.1^\circ \times 0.1^\circ$ grid was calculated as the product of the grid concentration and population density divided by global average population density over land area⁴².

Distribution of Individual Genetic Susceptibility. To characterize genetic susceptibility of the global population, data on genetic polymorphisms of the major human races were collected from the literature. Relative risks for lung cancer of different genotypes from 16 polymorphisms were calculated for major ethnic groups individually. Using the Monte Carlo simulation, 100,000 hypothetical persons for each ethnic group were generated with their genotypes randomly chosen based on genotype frequencies. The relative risk of lung cancer for an individual with the generated gene sequence was calculated as the product of the relative risks of all his/her genotypes. The distribution of the calculated relative risks of the 100,000 hypothetical persons (*GeneSus*) represented the overall distribution of genetic susceptibility of the specified ethnic group. See the Supplementary Methods for details.



Incremental Lifetime Cancer Risk. *ILCR* is commonly calculated as the product of lifetime average daily doses (*LADD*) and cancer slope factor (*CSF*)²⁴. In this study, a term of overall susceptibility (*SUS*) was introduced to describe the effect of individual susceptibility on the *ILCR* induced by exposure to PAHs in ambient air²⁹. *ILCR* was calculated as follows:

$$ILCR = CSF \times LADD \times SUS$$

$$= \sum_a \left[CSF \times \frac{(C \times IR_{a,g,r} \times y_{a,g,r})}{BW_{a,g,r} \times LE} \times (GeneSus_e \times EAF_e \times ASF_e) \right] \quad (3)$$

where a, g, e, and r are subscripts representing age, gender, ethnicity, and geographic region, respectively, suggesting that *LADD* is age, gender, and region dependent and *SUS* is ethnicity and region dependent; *C* (mg/m³) is BaP exposure concentration; *IR* (m³/day) is the inhalation rate; *y* (year) is the exposure duration; *BW* (kg) is body weight; *LE* (70 years) is the average life expectancy of the global population²⁷; *SUS* is the product of *GeneSus*, *EAF*, and *ASF*, which are genetic susceptibility, ethnicity-adjusted factor, and age-sensitivity factor, respectively. *GeneSus*, *EAF*, and *ASF* are all dimensionless and either ethnicity or region dependent.

In this study, a *CSF* of 26.6 kg(body weight)·day/mg for BaP was adopted as the maximum likelihood estimate based on epidemiological data from studies on coke-oven workers, using a multistage type model³¹. Use of BaP as an indicator of PAH exposure risk has been thoroughly investigated previously, and it was found that BaP can serve as an indicator for carcinogenic activity of PAH compounds at various sites⁴³.

The uncertainty of the *CSF* was derived according to the calculated disparity from different epidemiological studies²⁹. The resulting standard deviation of log-transformed *CSF* is 0.38. *C* is downscaled BaP concentrations in 2007 at 0.1° × 0.1° resolution. Derivation of the distributions of country- and age-specific *IR* and *BW* is described in the Supplementary Methods. *EAF* was calculated based on the lung cancer incidences for individual ethnicities reported by the United States Cancer Statistics⁴⁴, excluding the influence of smoking. *ASF* values of 10, 2, and 1 were used for the age groups of <2, 2–16, and >16 years, respectively²³. Male to female ratios were assumed to be 1 : 1 for all countries. Proportions of ethnic groups of each country were derived from a report of the Central Intelligence Agency of the United States⁴⁵.

To quantify population *ILCR* and uncertainty for all countries, *ILCR* values were calculated for 7 billion individuals in the world. The exposure concentrations were derived based on the spatial distributions of both BaP concentration and population⁴². Gender, ethnicity, genotypes, *IR*, and *BW* of an individual were randomly drawn from their distributions in given counties. Details can be found in the Supplementary Methods.

- United States Environmental Protection Agency. *Provisional guidance for quantitative risk assessment of polycyclic aromatic hydrocarbons* <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=49732> (accessed 6 April 2014).
- Boffetta, P., Jourenkova, N. & Gustavsson, P. Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons. *Cancer Causes Control* **8**, 444–472 (1997).
- Chen, S. C. & Liao, C. M. Health risk assessment on human exposed to environmental polycyclic aromatic hydrocarbons pollution sources. *Sci. Total Environ.* **366**, 112–123 (2006).
- Armstrong, B., Hutchinson, E., Unwin, J. & Fletcher, T. Lung cancer risk after exposure to polycyclic aromatic hydrocarbons: a review and meta-analysis. *Environ. Health Perspect.* **112**, 970–978 (2004).
- Brüske-Hohlfeld, I. *et al.* Occupational lung cancer risk for men in Germany: results from a pooled case-control study. *Am. J. Epidemiol.* **151**, 384–395 (2000).
- International Agency for Research on Cancer. *Agents classified by the IARC monographs, volumes 1–109* <http://monographs.iarc.fr/ENG/Classification/index.php> (accessed 6 April 2014).
- Ravindra, K., Sokhi, R. & Grieken, R. V. Atmospheric polycyclic aromatic hydrocarbons: source attribution, emission factors and regulation. *Atmos. Environ.* **42**, 2895–2921 (2008).
- Shen, H. Z. *et al.* Global atmospheric emissions of polycyclic aromatic hydrocarbons from 1960 to 2008 and future predictions. *Environ. Sci. Technol.* **47**, 6415–6424 (2013).
- Lunde, G. & Bjorseth, A. Polycyclic aromatic hydrocarbons in long-range transported aerosols. *Nature* **268**, 518–519 (1977).
- Lammel, G., Sehili, A. M., Bond, T. C., Feichter, J. & Grassl, H. Gas/particle partitioning and global distribution of polycyclic aromatic hydrocarbons – a modelling approach. *Chemosphere* **76**, 98–106 (2009).
- Zhang, Y., Tao, S., Ma, J. & Simonich, S. Transpacific transport of benzo[a]pyrene emitted from Asia. *Atmos. Chem. Phys.* **11**, 11993–12066 (2011).
- Friedman, C. L. & Selin, N. E. Long-range atmospheric transport of polycyclic aromatic hydrocarbons: a global 3-D model analysis including evaluation of Arctic sources. *Environ. Sci. Technol.* **46**, 9501–9510 (2013).
- Wang, R. *et al.* Exposure to ambient black carbon derived from a unique inventory and high-resolution model. *Proc. Natl. Acad. Sci. USA* doi:10.1073/pnas.1318763111 (2014).
- Perera, F. P. Environment and cancer: who are susceptible? *Science* **278**, 1068–1073 (1997).
- Mumford, J. L., Li, X., Hu, F., Lu, X. B. & Chuang, J. C. Human exposure and dosimetry of polycyclic aromatic hydrocarbons in urine from Xuan Wei, China with high lung cancer mortality associated with exposure to unvented coal smoke. *Carcinogenesis* **16**, 3031–3036 (1995).
- Zhang, Y. X., Tao, S., Shen, H. Z. & Ma, J. M. Inhalation exposure to ambient polycyclic aromatic hydrocarbons and lung cancer risk of Chinese population. *Proc. Natl. Acad. Sci. USA* **106**, 21063–21067 (2009).
- Emmons, L. K. *et al.* Description and evaluation of the model for ozone and related chemical tracers, version 4 (MOZART-4). *Geosci. Model. Dev.* **3**, 43–67 (2010).
- Task Force on Hemispheric Transport of air pollution. *Theme 2: source apportionment and source/receptor analysis* <http://www.htap.org/> (accessed 6 April 2014).
- ESRI Inc. *ArcGIS for Desktop* <http://www.esri.com> (accessed 6 June 2014).
- Microsoft Corporation. *All Office products* <http://office.microsoft.com> (accessed 6 June 2014).
- European Union Directive. Directive 2004/107/EC of the European Parliament and of the Council. *Official Journal of the European Union* **L23**, 3–16 http://eur-lex.europa.eu/LexUriServ/site/en/oj/2005/l_023/l_02320050126en00030016.pdf (accessed 15 December 2013).
- Ministry of Environmental Protection of the People's Republic of China. *Ambient air quality standards (GB 3095–2012)* <http://www.zhb.gov.cn/> (accessed 6 April 2014).
- California Environmental Protection Agency. *Technical support document for cancer potency factors: methodologies for derivation, listing of available values, and adjustments to allow for early life stage exposures* http://oehha.ca.gov/air/hot_spots/2009/TSDCancerPotency.pdf (accessed 6 April 2014).
- Health Canada. *Federal contaminated site risk assessment in Canada, part I: guidance on human health preliminary quantitative risk assessment (PQRA), Version 2.0* <http://www.healthcanada.gc.ca/> (accessed 6 April 2014).
- Ezzati, M. & Lopez, A. D. Estimates of global mortality attributable to smoking in 2000. *Lancet* **362**, 847–852 (2003).
- Ferlay, J. *et al.* GLOBOCAN 2012 v1.0. *Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11* [Internet] Lyon, France: International Agency for Research on Cancer. <http://globocan.iarc.fr>. (accessed 6 April 2014).
- World Health Organization. *Global Health Observatory Data Repository* <http://apps.who.int/gho/data/node.main> (accessed 6 April 2014).
- Liu, B. *et al.* Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. *BMJ* **317**, 1411–1422 (1998).
- Boström, C. E. *et al.* Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environ. Health Perspect.* **110**, 451–488 (2002).
- Mastrangelo, G., Fadda, E. & Marzia, V. Polycyclic aromatic hydrocarbons and cancer in man. *Environ. Health Perspect.* **104**, 1166–1170 (1996).
- United States Environmental Protection Agency. *Carcinogen Assessment of Coke Oven Emissions, Final Report* <http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=47897> (accessed 6 April 2014).
- Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project. *B. Am. Meteorol. Soc.* **77**, 437–471 (1996).
- Lohmann, R. & Lammel, G. Adsorptive and absorptive contributions to the gas-particle partitioning of polycyclic aromatic hydrocarbons: state of knowledge and recommended parametrization for modeling. *Environ. Sci. Technol.* **38**, 3793–3803 (2004).
- Esteve, W., Budzinski, H. & Villenave, E. Relative rate constants for the heterogeneous reactions of NO₂ and OH radicals with polycyclic aromatic hydrocarbons adsorbed on carbonaceous particles. Part 2: PAHs adsorbed on diesel particulate exhaust SRM 1650a. *Atmos. Environ.* **40**, 201–211 (2006).
- Kahan, T. F., Kwamena, N.-O. A. & Donaldson, D. J. Heterogeneous ozonation kinetics of polycyclic aromatic hydrocarbons on organic films. *Atmos. Environ.* **40**, 3448–3459 (2006).
- Bond, T. C. *et al.* A technology-based global inventory of black and organic carbon emissions from combustion. *J. Geophys. Res.* **109**, D14203 (2004).
- Zhu, D. *et al.* Temporal and spatial trends of residential energy consumption and air pollutant emissions in China. *Appl. Energy* **106**, 17–24 (2013).
- van der Werf, G. R. *et al.* Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmos. Chem. Phys.* **10**, 11707–11735 (2010).
- Shen, Z. Y. Long period average concentration of single continuous source gas effluent at one and multi-receiving point. *Environmental Pollution Control* **2**, 37–41 (1987) (in Chinese).
- Sørensen, J. H. Sensitivity of the derma long-range Gaussian dispersion model to meteorological input and diffusion parameters. *Atmos. Environ.* **32**, 4195–4206 (1998).
- Brauer, M. *et al.* Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. *Environ. Sci. Technol.* **46**, 652–660 (2012).
- Oak Ridge National Laboratory. *LandScan Global Population 2007 Database* <http://www.ornl.gov/sci/landscan/> (accessed 6 April 2014).
- Expert Panel on Air Quality Standards. (1999) *Choice of a marker compound for polycyclic aromatic hydrocarbon* <http://web.archive.nationalarchives.gov.uk/20060715141954/http://www.defra.gov.uk/environment/airquality/aqs/poly/6.htm> (accessed 6 April 2014).



44. National Program of Cancer Registries. *United States Cancer Statistics (USCS)* http://www.cdc.gov/cancer/npcr/uscs/download_data.htm (accessed 6 April 2014).
45. Central Intelligence Agency. *The World Factbook* <https://www.cia.gov/library/publications/the-world-factbook/> (accessed 6 April 2014).

Acknowledgments

We thank Louisa K. Emmons for suggestions on Mozart-4 performance; Yanxu Zhang and Rong Wang for advices on module development and validation. We also acknowledge the discussions with Zhaobin Zhang, Yilong Wang, and Jiachen Zhang. This study was supported by the National Natural Science Foundation of China (Grant 41390240 and 41130754).

Author contributions

T.S. and H.S. conceived the study. H.S. prepared the BaP emission inventory. H.S., J.L., Y.H., W.Li. and H.C. developed the PAH module and incorporated it into MOZART-4. Y.H., W.Li. and H.C. collected the observation data for model validation. H.S. and J.L. achieved the model debugging and performed the long-term and short-term simulations. T.S. and H.S. carried out the downscaling method and conducted the exposure assessment. X.W. and W. Liu suggested the dose-response relationship. Y.Z., S.S., N.L. and Y.X. contributed to the

collection of the information on genetic polymorphisms together with individual differences of other factors. H.S., Y.C. and B.L. conducted the Monte Carlo simulation of individual susceptibility, based on which H.S. and S.T. finished the risk assessment through model outputs by source type and region. H.S. and S.T. drafted the paper. All other authors contributed to the interpretation of the results and to the text.

Additional information

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

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

How to cite this article: Shen, H. *et al.* Global lung cancer risk from PAH exposure highly depends on emission sources and individual susceptibility. *Sci. Rep.* 4, 6561; DOI:10.1038/srep06561 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>