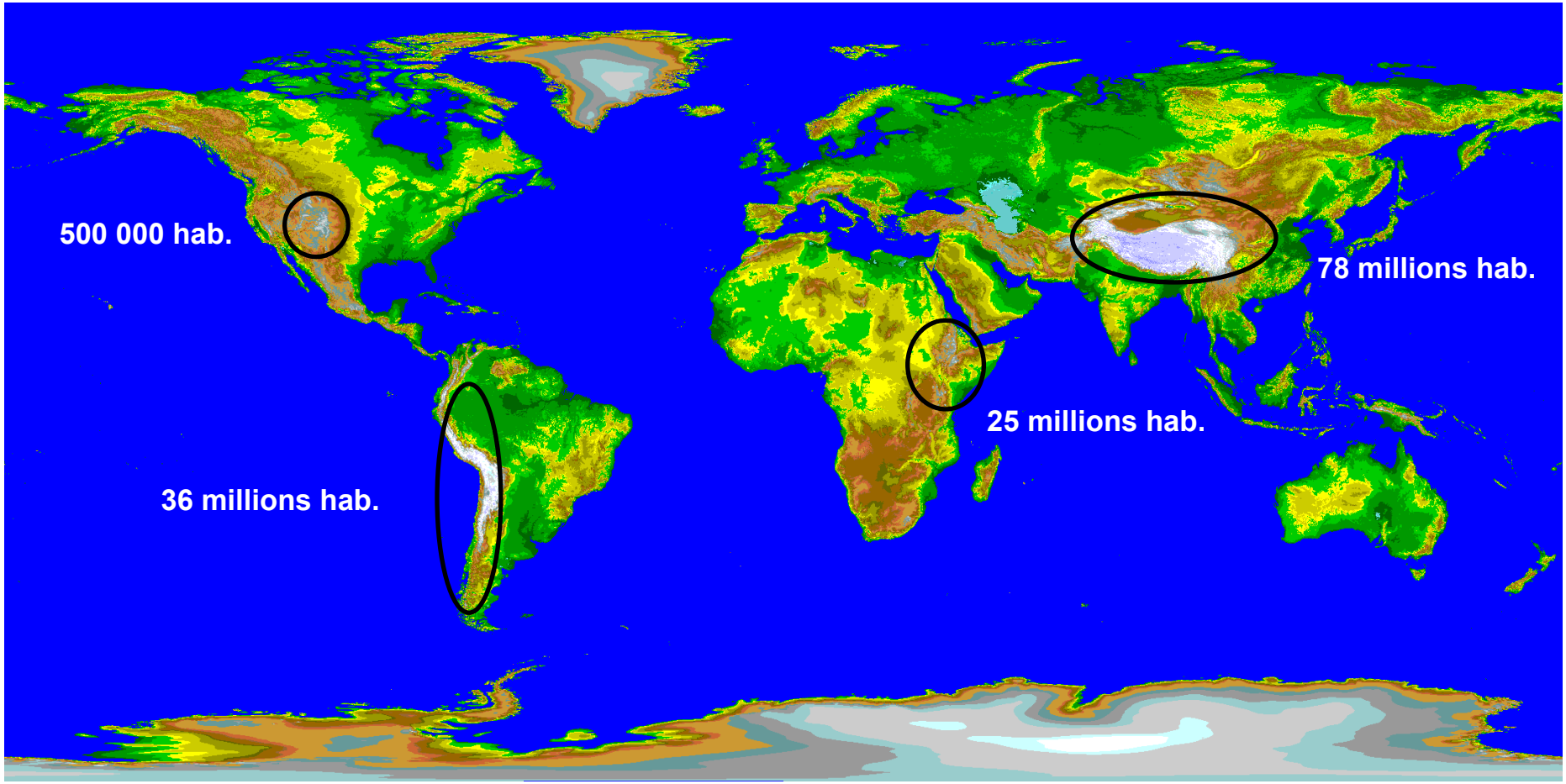
Complex patterns of natural selection into the genetic architecture of chronic mountain sickness

Jean-Paul Richalet, Steven Gazal, Jose Espinoza, Frédéric Austerlitz, Jose Luis Macarlupu, Fabiola León-Velarde, Francisco Villafuerte, Dominique Marchant, Laurent Gouya, Olivier Hermine.

Université Paris 13, Sorbonne Paris Cité, EA2363 Hypoxie et Poumon, Bobigny, Laboratoire d'excellence GrEx, France
Department of Epidemiology, Harvard School of Public Health, Boston, USA
INSERM U1149, Hème, fer et pathologies inflammatoires, Paris,
INSERM U946, Variabilité Génétique et Maladies Humaines, Paris
UMR CNRS 7206 Eco-Anthropologie et Ethnobiologie, Musée de l'Homme, Paris
Universidad Peruana Cayetano Heredia, Laboratories for Research and Development, Molecular Biotechnology and High Altitude Units, Lima, Peru

5th International Leh Symposium Gangtok, Sikkim, 2018

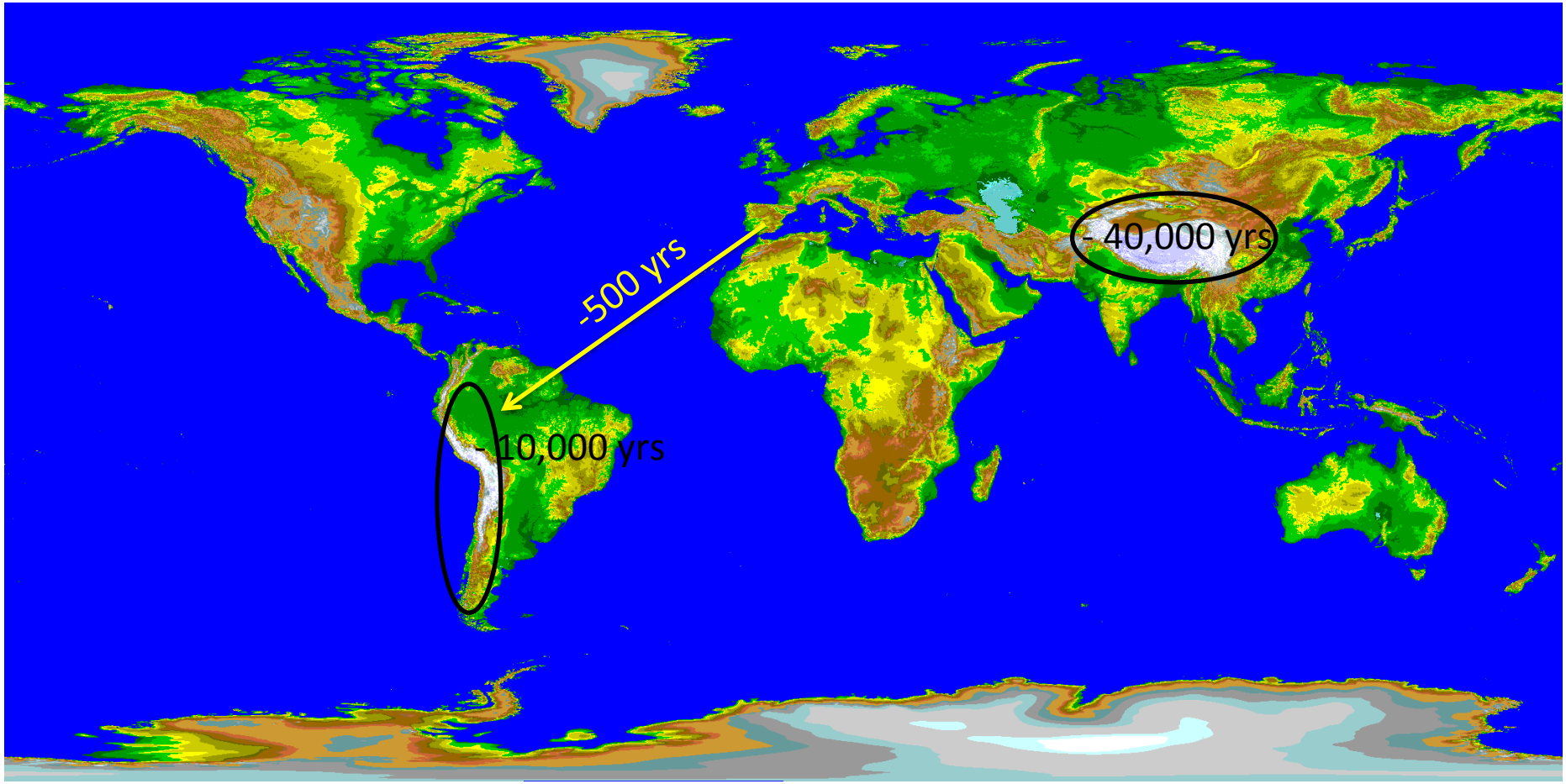
Permanent life at high altitude



140 million persons live permanently above 2500m of altitude

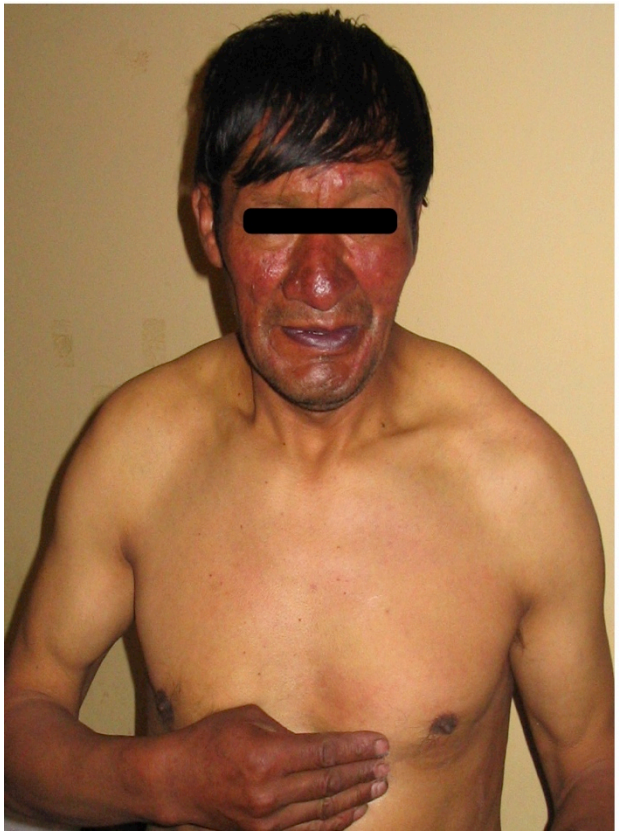
Moore LG et al. 1998 and WHO

Permanent life at high altitude



Andes: admixture of European genome….

High Altitude Excessive Erythrocytosis (Monge's disease)



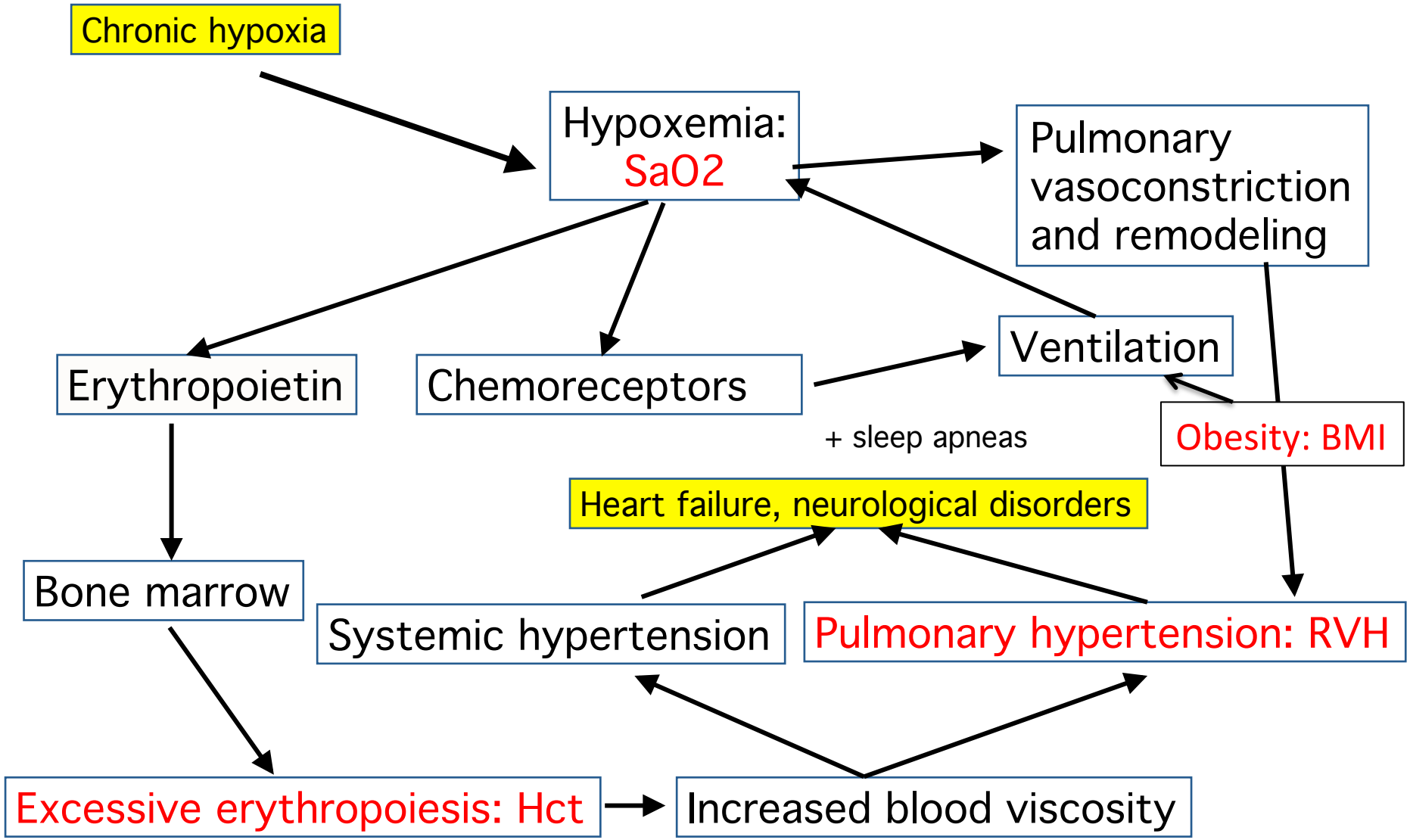
Headache, insomnia, fatigue, anorexia, cyanosis, tinnitus, ...

- Monge's disease is an excessive erythrocytosis:
  - males : Hb > 21 g/dl
  - females: Hb > 19 g/dl
- In some cases associated with pulmonary hypertension.
- Mainly males and post-menopausal women
- Prevalence: 5 to 15%

Hurtado, 1964; Spielvogel et al., 1981

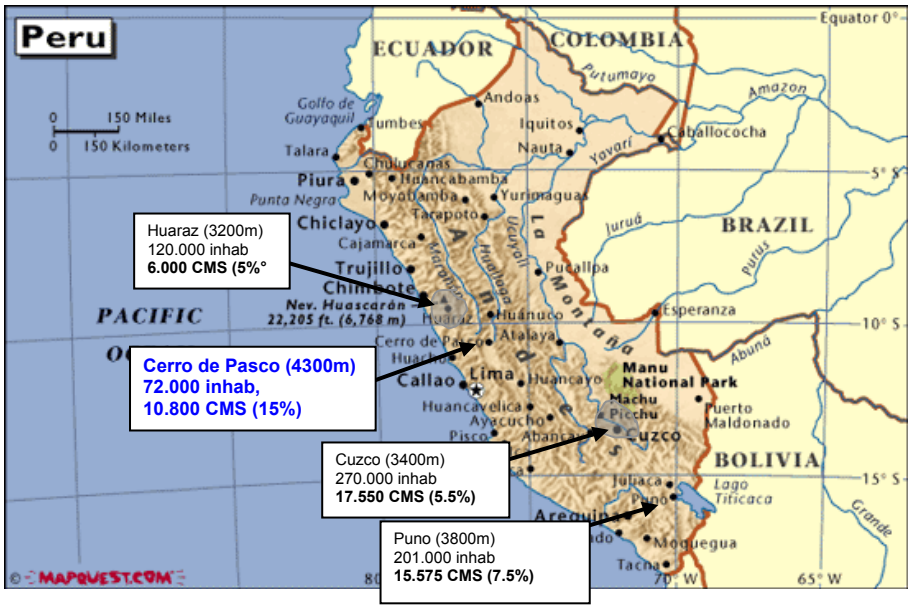
Monge C.,1966; Monge et al., 1989
Leon Velarde, 1993

Pathophysiology of Monge's disease



Monge GWAS study 2010-2015

Objective: find genetic determinants of Monge's disease



Monge GWAS study 2010-2014

- Subjects
  - From several trips to Cerro de Pasco (4300m, Peru), 340 patients were sampled
    - 184 patients (Hct ≥ 63%)
    - 156 controls (Hct ≤ 57%)
  - For various reasons (loss of samples, degraded DNA, etc..), 244 subjects were finally selected:
    - 166 patients
    - 146 controls

Quality Control (QC) of GWAS data

- 4,301,332 SNPs available
  - 12,662 SNPs are not on chromosomes 1 to 22, X or Y
  - 100,057 SNPs have > 5% of missing genotypes
  - 7,507 SNPs have missing genotype rates different between cases and controls ($p<0.05$)
  - 235 SNPs departed from HW equilibrium in controls ($p<10^{-5}$)
  - 2,858,934 SNPs have a minor allele frequency < 5% in controls (including 1,952,249 monomorphic SNPs)
  - 406,554 SNPs are in strong linkage disequilibrium (LD) with another SNP ($r^2 > 0.99$)
- **915,383 SNPs remaining**

Clinical and physiological evaluation

- History, BMI
- EKG (signs of pulmonary hypertension), systemic blood pressure
- Pulmonary function test, pulse oximetry: SaO_2

Statistical analysis of Monge GWAS dataset was performed in 3 steps

1. Preliminary association analysis

To have a quick overview of the data, and to check if common variants are highly associated with the phenotype (Odd Ratio > 4).

2. Population genetic description

To define population structure and the amount of relationships/inbreeding present in the genetic data.

- Our population is an admixture of European/Spanish and Native American populations, but its structure is unknown.
- For this reason we will use the HGDP-CEPH panel (Cann et al., Science, 2002), a dataset of 1000 individuals (including Native Americans) from 50 different populations from 6 geographic regions in all inhabited continents genotyped for 600,000 SNPs, to first perform a principal component analysis and define the individual structure of each individual.
- Finally, we estimated relationships between individuals and inbreeding within individual from genotype data.

3. Final association analysis

Final association analysis has been performed with a linear mixed model, which takes into account both admixture and relatedness, such as performed by Moltke et al. (Nature, 2014).

Clinical and physiological data (1)

Variable	CMS (n=143)	CTRL (n=124)	P
Age (years)	46.8±13.4	43.0 ± 13.1	NS
Body weight (kg)	68.8 ± 9.2	64.9 ± 8.7	<0.001
Height (cm)	163 ± 6	163 ± 6	ns
Body Mass Index (kg/m^2)	26.1 ± 3.3	24.24 ± 2.7	<0.001
Hematocrit (%)	67.2 ± 3.8	52.0 ± 3.2	<0.001
CMS clinical score	6.8 ± 3.6	2.2 ± 2.1	<0.001
SaO_2 (%)	85.3 ± 5.2	89.7 ± 4.8	<0.001
Heart rate (b/min)	72.1 ± 3	67.6± 9.2	<0.001
Systolic BP (mmHg)	116.3 ± 15.2	115.4 ± 13.6	ns
Diastolic BP (mmHg)	76.9 ± 10.3	76.0 ± 10.1	ns
RV hypertrophy (EKG)	120 (72%)	61 (42%)	<0.001
Quality of life score	68.8±15.1	70.5±12.7	ns

Multinominal logistic regression

- Variables associated with non-polycythemia:

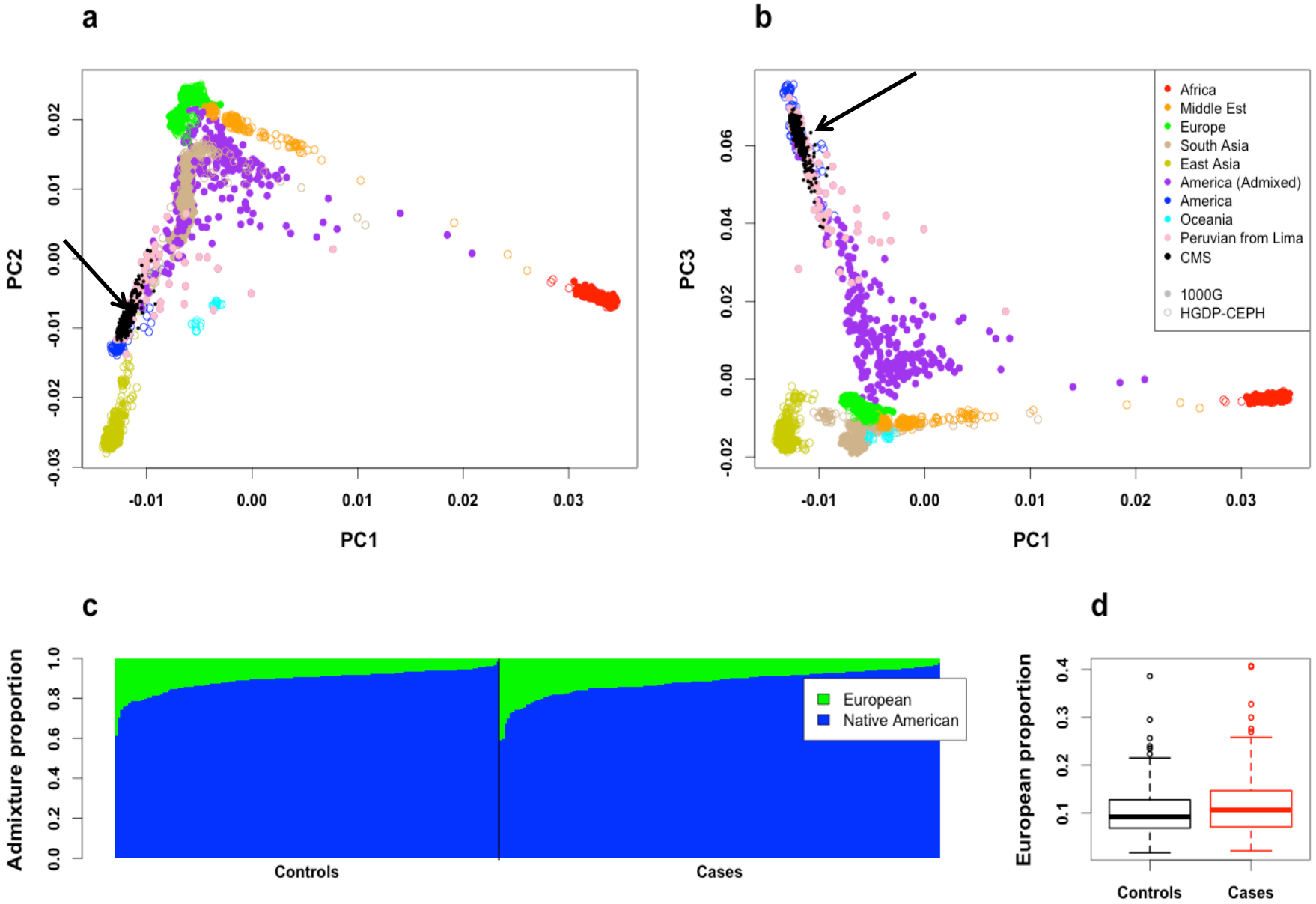
  SaO_2: OR= 1.18 (1.10-1.27) p<0.001

  BMI : OR=0.83 (0.73-0.95) p=0.006

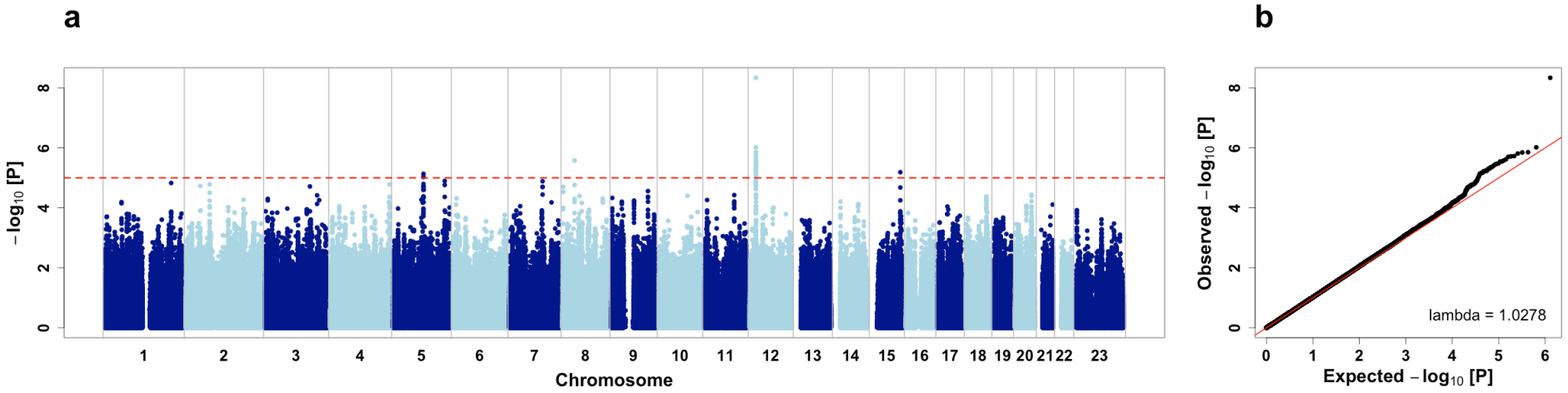
  CMS clinical score: OR=0.65 (0.57-0.75) p<0.001

Genotypic population structure

Projection of the 312 subjects (in black) on panels of HGDP-CEPH and 1000 genomes.

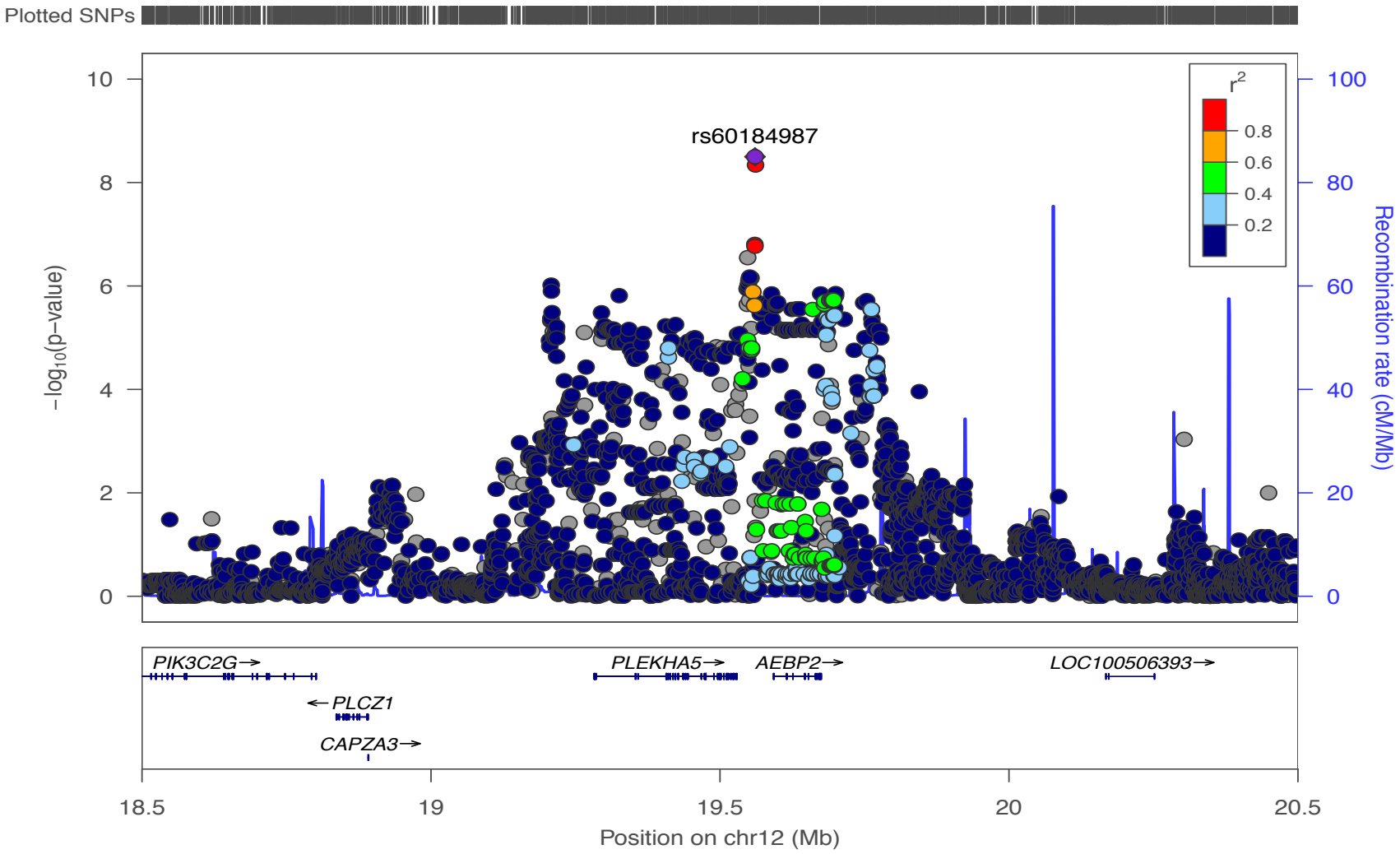


Association (1) - Manhattan plot



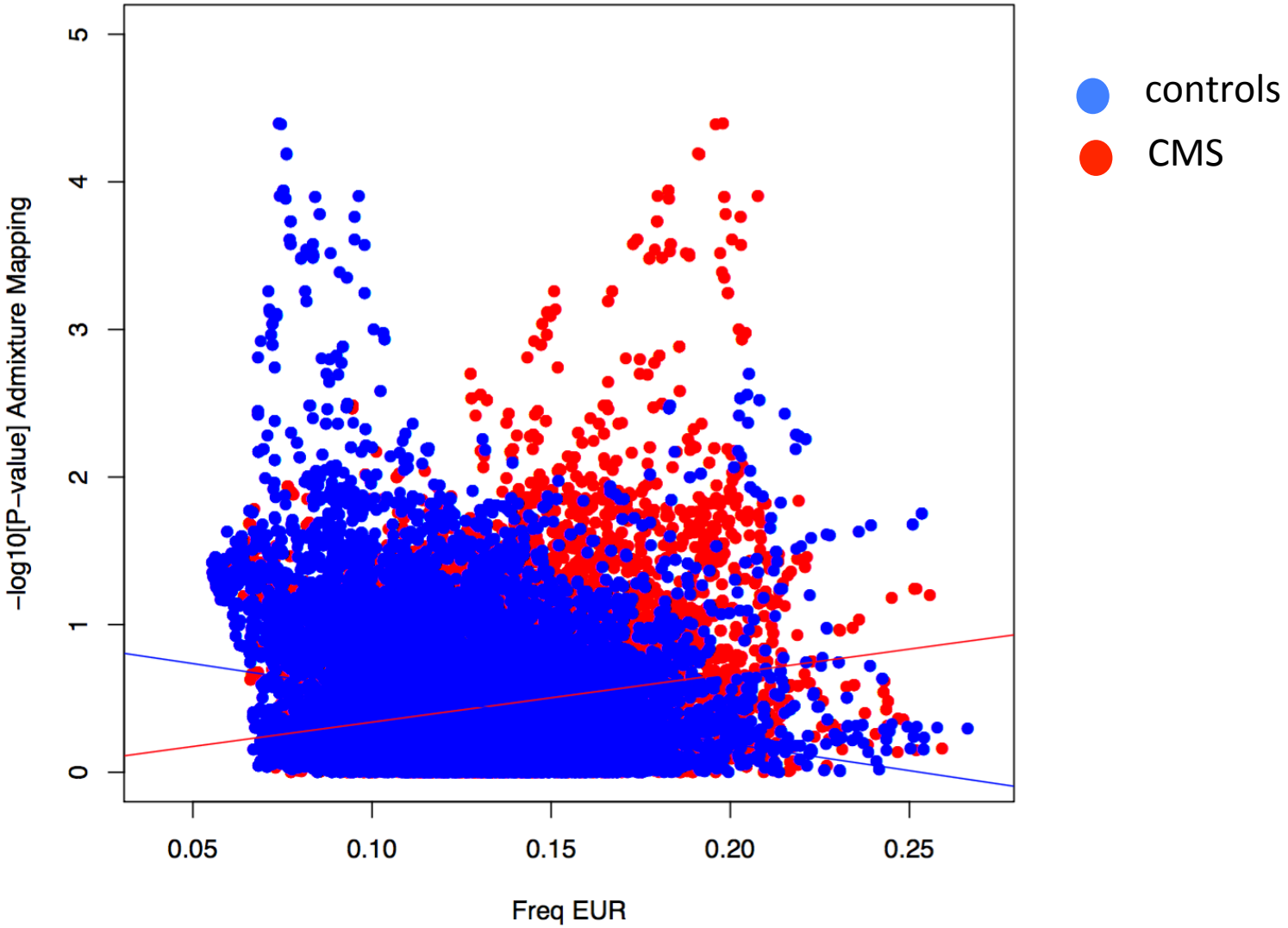
Variant	Chr	Position (hg19)	Alleles	OR [95% CI]	*P* GWAS	Nearest genes	Candidate Gene
rs7304081	12	19,561,543	**A**/C	2.52 [1.85; 3.48]	**4.58×10^{-9}**	*PLEKHA5, AEBP2*	*AEBP2*
rs75810402	5	96,747,224	A/**G**	0.28 [0.13; 0.50]	7.37×10^{-6}	*RIOK2, LINC01340*	CAST
rs7832232	8	38,469,303	**G**/A	2.16 [1.56; 3.04]	2.63×10^{-6}	*RNF5P1, TACC1*	-

Candidate region of chromosome 12.

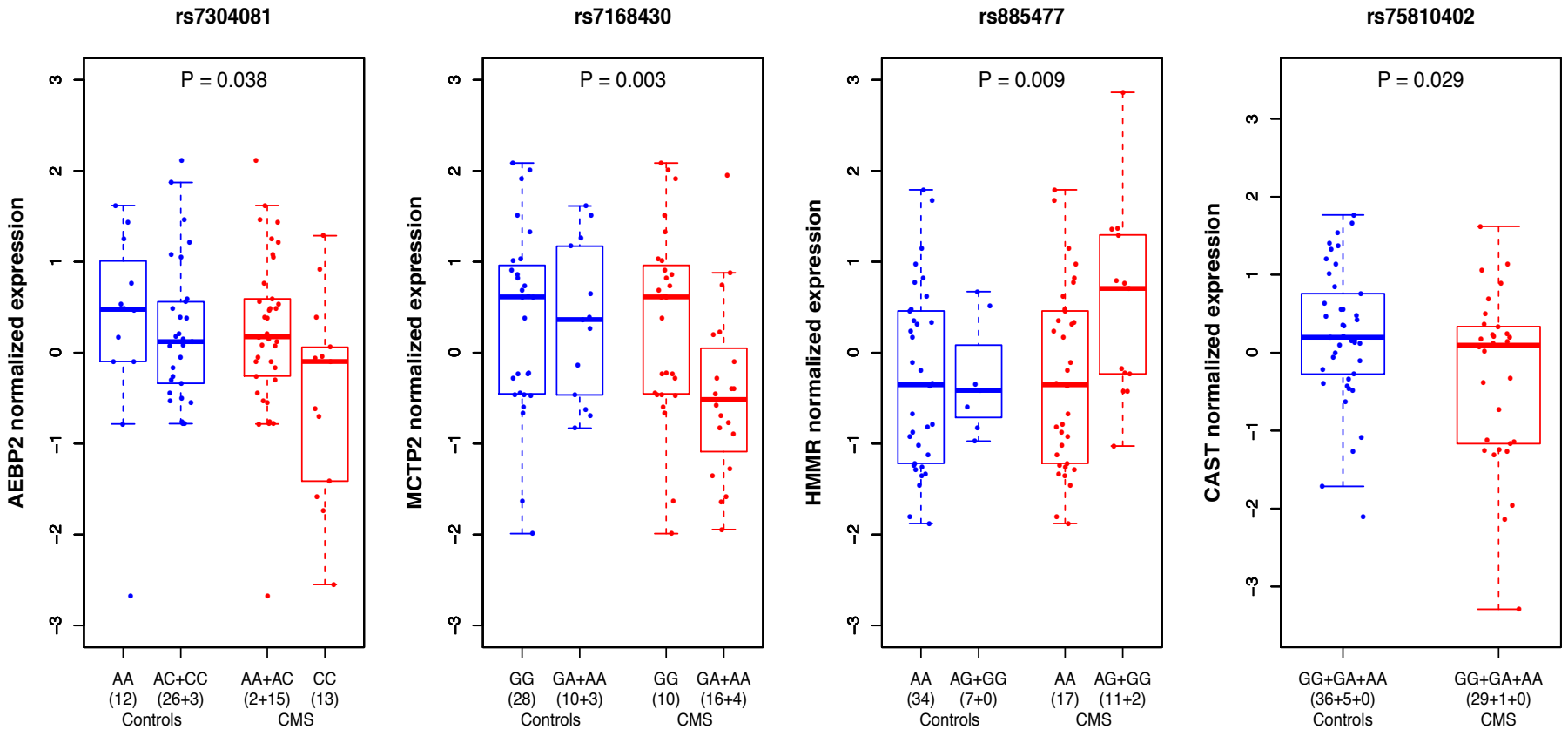


P values across the candidate regions for the imputed SNPs plotted using the LocusZoom software (Pruim et al., 2010). The SNP with the lowest *P* value is plotted as a purple diamond; other data points are colored according to their r^2 in the American population of 1000 Genomes.

Role of admixture in CMS
Local excess of european genome (ns)



Expression analyses around candidate variants (ARN quantification from whole blood)



CMS (n=30 in red); Controls (n=41 in blue)

Association between expression of "literature" genes and CMS status

Candidate Loci	Gene name	Effect size (s.e.)	*P* value
Literature	EPAS1	-0.80 (0.42)	0.060
	EGLN1	-1.37 (0.72)	0.056
	EDNRA	0.06 (0.18)	0.746
	SENP1*	***-2.59 (0.98)***	***0.008***
	ANP32D	0.11 (0.19)	0.556
	PRKAA1	-0.56 (0.57)	0.328
	EDNRB	-0.06 (0.47)	0.904
	ATM*	***-1.45 (0.71)***	***0.041***
	VEGFA*	***-1.68 (0.73)***	***0.021***
	PDP2	-0.83 (0.51)	0.101

In conclusion

- Chronic mountain sickness is associated with low SaO_2 and high BMI: should be taken into account in genetic studies
- Some genes seem to be associated with high Hct:
  - AEBP2 on chromosome 12 ($P=4.6\ 10^{-9}$)
  - CAST, HMMR on chromosome 5
  - MCTP2 on chromosome 15
- European admixture may have played in role in de-adaptation to life at high altitude
- No interesting signals for EPAS1/HIF2A or EGLN1/PHD2 genes in association with CMS
- None of candidate loci were under strong natural selection

  (CMS happens late in life, after the reproduction period)
- Expression analysis confirmed the potential role of candidate genes