Table S1. STSs Used in Detecting and Characterizing Deletions Involving *AZFc*

STS Identifier	GenBank Accession [a]
STSs used to detect deletions in first-stage screen:	
sY1191 [b]	G73809
sY1291	G72340
STSs used to characterize deletions in second stage:	
sY14	G38356
sY254	G38349
sY1192	G67166
sY1189	GF102061
sY1201	G67170
Additional STSs used to characterize unusual deletions:	
sY627	G67175
sY1196	G67167
sY1197	G67168
sY160	G38343
sY159	G38354
sY1273	G75500

[a] See GenBank entries for reaction conditions.

[b] For the first stage of screening, we assayed sY1191 using TaqMan with probe 6FAM 5'-ACA CGT AAT AGG CTT TAC ATT A-3' with a "Minor groove binder/Non-fluorescent quencher" at the 3' end (Applied Biosystems). Primers were as specified in GenBank G73809. We thermal cycled on an Applied Biosystems Geneamp PCR System 3700 and assessed presence or absence of amplification with an end-point read on an Applied Biosystems 7900HT real time PCR machine. The reaction conditions were:

Master Mix, Applied Biosystems P/N 4324018	2.5 ul
Primer 1 @ 20 uM	0.25 ul
Primer 2 @ 20 uM	0.25 ul
TaqMan probe @ 250 nM	0.5 ul
DNA @ 5ng / ul	1.5 ul

Thermal cycling:

1. 95° 10 min
2. 94° 1 min
3. 61° 1 min
4. 72° 2 min

Repeat steps 2-4 34 more times

5. 72° 10 min

We used the "Allele Y Rn" collumn of the 7900HT output to determine whether a sample possibly bore a deletion. Each plate contained multiple wells with negative controls, including DNA from men with known deletions at sY1191, DNA from women, and water. We set a cut-off of mean(NC) + 8 * sd(NC), where mean(NC) denotes the mean "Allele Y Rn" of the known negative controls and sd(NC) represents the corresponding standard deviation. Samples with "Allele Y Rn" at or below the cutoff were carried forward to the second stage of testing as possibly lacking the site detected by sY1191.

Table S2. STS Results for Unusual Deletions Overlapping sY1192 or sY1291

Sample ID	Study Population	Self-Reported Ethnicity	Y Haplogroup [b]	STS [a] sY14	sY142	sY627	sY1196	sY1197	sY1192	sY1191	sY1291	sY1189	sY254	sY1201	sY159	sY160	sY1273
I8GHS	Poland	White	R1a	■	■	■	–	Lg [c]	–	–	–	–	■	■	■	■	
PIBMR [d]	Poland	White	F*xK	■	■	■	■	■	–	–	■	■	■	–	–	–	–
STSs patterns for recurrent deletions for comparison:																	
gr/gr				■	■	■	■	■	■	■	–	–	■	■	■	■	■
b2/b3				■	■	■	■	■	–	–	■	■	■	■	■	■	■
b1/b3				■	■	■	–	–	–	–	–	–	■	■	■	■	■
b2/b4				■	■	■	■	■	–	–	–	–	–	■	■	■	■

"-" indicates consistent absence of PCR product; black-filled cell indicates presence of PCR product; gray cell indicates STS not tested but presumed to be present.

[a] Refer to Figure 1 in the main text and Table S1; in addition, sY1196 and sY1197 occur at the edges of the center of the palindrome at the right end of Figure 1B (palindrome P3); sY159 and sY160 occur in the distal-Yq heterochromatin; sY1273 detects the presence of the boundary between the MSY (male specific region of the Y chromosome) and PAR2 (the second pseudoautosomal region, on distal Yq, Figure 1).

[b] Y Haplogroup according to nomenclature of Karafet et al. (2008) Genome Res. *18* , 830-838; R1a assayed as sY207-C, SRY10831-A; F*xK as M89-T, M9-C.

[c] "Lg" indicates a larger-than-expected PCR product.

[d] The PIBMR Y chromosome lacks both the distal-Yq heterochromatin (sY159 and sY160) and the boundary between the MSY and PAR2 (Figure 1, detected by sY1273). This STS pattern likely arises from an isodicentric chromosome, as reported in Lange et al. (2009) Cell, *138* , 855-869.

Table S3. b2/b4 Deletions

Sample ID	Study Population	Self-Reported Ethnicity	Y Haplogroup [a]	Genotypes Used to Determine Haplogroup [a]
1YRU5	Poland	White	R1a	M89-T, M9-G, M207-C, SRY10831-A
2KR7Z	Poland	White	R*xR1a	M89-T, M9-G, M207-C, SRY10831-G
5SAAI	USA	White	R*xR1a	M89-T, M9-G, M207-C, SRY10831-G
5WQDU	USA	Hispanic	R*xR1a	M89-T, M207-C, SRY10831-G
81MEK	USA	White	E	M89-C, M9-C, M203-C, M96-C
KH9Q8	USA	White	R*xR1a	M89-T, M9-G, M207-C, SRY10831-G
KYGGV	USA	White	R*xR1a	M89-T, M9-G, M207-C, SRY10831-G
RMHV9	USA	White	R*xR1a	M89-T, M207-C, SRY10831-G
RUXF9	USA	Hispanic	K*xR	M89-T, M9-G, M207-T

[a] See Karafet et al. (2008) Genome Res. *18,* 830-838.

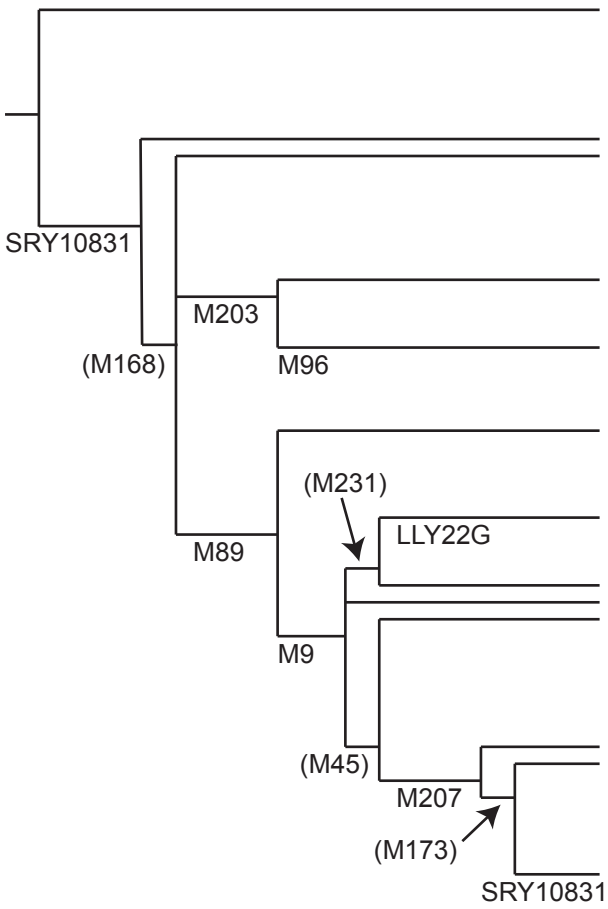
Table S4. Prevalence of b2/b3 Deletions Not in Branch N1
Variation across populations is not significant (p > 0.15 by Fisher's exact test, two-sided, on the first two rows of this table).

	India *n* (%)	Poland *n* (%)	Tunisia *n* (%)	US *n* (%)	Vietnam *n* (%)	All Populations *n* (%)
b2/b3-deleted and not in branch N1	2 (0.50)	21 (0.46)	3 (0.52)	45 [a] (0.30)	1 (0.93)	72 (0.35)
Non-b2/b3-deleted [b]	402 (99.50)	4,566 (99.54)	575 (99.48)	15,003 (99.70)	106 (99.07)	20,652 (99.65)
Total Samples	404	4,587	578	15,048	107	20,724

[a] Includes two b2/b3-deleted chromosomes that we determined are not in branch N1, although we could not fully determine their haplogroups; these chromosomes are not counted in Table 2.

[b] Presumed non-branch-N1, neglecting the very small proportion of branch-N1 chromosomes that may not be b2/b3-deleted.

Table S5. Information on Haplogroup Definitions and Nomenclature. This table provides more information on Figure 3B. In the left part of this table, polymorphisms in parentheses, e.g. "(M168)", were not typed as part of this study, but are provided for orientation.



Haplogroup	Mono-phyletic?	Notes
Y*xBF	N	Ancestral (A) at SRY10831 and M207. Chromosomes in this haplogroup are likely to be in branch A.
BC*xDEF	N	Derived (G) at SRY10831 and ancestral at M203 and M89. Chromosomes in this haplogroup may be ancestral or derived for M168 (not typed). Those that are ancestral at M168 are likely to be in branch B, while those that at derived at M168 are likely to be in branch C.
DE*xE	N	Derived at M203 and ancestral at M96. Chromosomes in this haplogroup are likely be in branch D.
E	Y	Derived at M96.
F*xK	N	Derived at M89 and ancestral at M9. Chromosomes in this haplogroup may be in finer haplogroups F*xGK, G, H, I or J.
N1	Y	Derived at LLY22g.
K*xN1,R	N	Derived at M9 and ancestral at LLY22G and M207. Chromosomes in this haplogroup are in one the finer haplogroups N*xN1, P*xR, or K*xNP.
R*xR1a	N	Derived at M207 and ancestral at SRY10831. Chromosomes in this haplogroup that are ancestral at M173 (not typed) are in haplogroup R*xR1. Chromosomes in this haplogroup that are derived at M173 are in haplogroup R1*xR1a.
R1a	Y	SRY10831 A and derived at M207.

Table S6. Counts of Y Chromosomes in Each Haplogroup by Deletion Class for Two Populations

Main text Figure 3A is based on the data and analysis in this table. As discussed in the text, branch N1 is excluded. Bold numbers in parentheses are standardized residuals [a] that are > 2 or < –2, which indicate major contributions to variation in haplogroup prevalence compared to non-deleted chromosomes. Values > 2 indicate an excess of deleted chromosomes in a given haplogroup; values < –2 indicate a deficit. It was not possible to completely determine the haplogroup of each deleted chromosome, so for some rows in this table the total number of deleted chromosomes is less than in Table 1.

		Y*xBF	BC*xDEF	DE*xE	E	F*xK	N1	K*xN1,R	R*xR1a	R1a	p vs. Non-Deleted [b]
Poland	Not Deleted [c]				12	49	(Excluded)		44	85	
	gr/gr				1	12 **(–2.3)**		1	12	89 **(2.9)**	1.1×10^{-7} *
	b2/b3		1 **(2.9)**		5 **(2.5)**	2		1 **(2.9)**	1	11	2.4×10^{-4} *
	b1/b3									1	1
US	Not Deleted [c]	2			72	129		18	216	41	
	gr/gr	1	2	2	45	101		22	117	20	0.019
	b2/b3		5 **(7.1)**		15 **(2.9)**	7		4	6 **(–2.9)**	6	1.4×10^{-9} *
	b1/b3	1				2			5	6 **(4.0)**	1.2×10^{-3} *

[a] Defined as $(n_o - n_e) / n_e^{0.5}$, where n_o is the observed count and n_e is the expected count (Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures: Fourth Edition*, Chapman and Hall/CRC, 2007).

[b] By Fisher's exact test, two-sided; * indicates significance at a Bonferroni corrected alpha of 0.0056.

[c] A subsample of non-deleted chromosomes.

Table S7. Prevalences of gr/gr and b1/b3 Deletions Among Men with Severe Spermatogenic Failure

	Azoospermic			Oligospermic (< 5e6 / ml)			
Source	**n**	**n gr/gr**	**n b1/b3**	**n**	**n gr/gr**	**n b1/b3**	**Notes**
(Repping et al., 2003)	249	9	0	198	6	1	Page laboratory patients with SSF from Repping et al., 2003
(Machev et al., 2004)	94	3	0	160	7	0	
(Ferlin et al., 2005)	73	3	0	193	10	0	
(Hucklenbroich et al., 2005)	61	2	0	133	6	1	Oligospermic is < 1e6 / ml
(Lynch et al., 2005)	197	7	3	307	12	1	Table II, first group
(Lynch et al., 2005)	119	6	0	329	15	0	Table III, second group
(de Carvalho et al., 2006)	49	NUA	0				
(Fernando et al., 2006)	79	NUA	0	15	NUA	0	Oligospermic is < 1e6 / ml
(Imken et al., 2007)	48	3	0				
(Navarro-Costa et al., 2007)	90	5	0				
(Wu et al., 2007)	164	NUA	0				
(Zhang et al., 2007)	216	NUA	0				
(Giachini et al., 2008)	72	0	0	213	10	0	Table 3 has gr/gr sperm counts
(Stouffs et al., 2008)	44	1	0	143	7	1	See methods, page 1194, left column
(Lu et al., 2009)	220	NUA	0				
(Ravel et al., 2009)	115	4	0	72	5	0	Oligospermic is < 1e6 / ml
(Visser et al. 2009)	68	5	NU	98	1	NU	This paper does not report b1/b3 deletions.
(Stahl et al., 2011)	989	43	NU	317	18	NU	b1b3 not reported; for gr/gr see table on page 235
This study	159	3	0	144	3	3	
Total for gr/gr	**2,378**	**94**		**2,307**	**100**		**Total n for gr/gr is 4685; total number of deletions is 194**
Total for b1/b3	**2,049**		**3**	**1,907**		**7**	**Total n for b1/b3 is 3956; total number of deletions is 10**

NU denotes “not used” in the calculation for that deletion type; see the corresponding Notes for explanation.

NUA denotes gr/gr data not used because Asian population.

NOTE: We excluded (Shahid et al., 2011), which found 16 b1/b3 deleted Y chromosomes in branch L1 (derived alleles at SNPs M20 and M176) among 236 men with azoospermia and 182 men with sperm counts < 20e6 / ml. No b1/b3 deletions were found among 240 “normospermic” controls.

References for Table S7

de Carvalho, C.M.B., Zuccherato, L.W., Fujisawa, M., Shirakawa, T., Ribeiro-dos-Santos, A.K.C., Santos, S.E.B., Pena, S.D.J., and Santos, F.R. (2006). Study of AZFc partial deletion gr/gr in fertile and infertile Japanese males. J Hum Genet *51*, 794-799.

Ferlin, A., Tessari, A., Ganz, F., Marchina, E., Barlati, S., Garolla, A., Engl, B., and Foresta, C. (2005). Association of partial *AZFc* region deletions with spermatogenic impairment and male infertility. J Med Genet *42*, 209-213.

Fernando, L., Gromoll, J., Weerasooriya, T.R., Nieschlag, E., and Simoni, M. (2006). Y-chromosomal. micro deletions and partial deletions of the Azoospermia Factor c (*AZFc*) region in normozoospermic, severe oligozoospermic and azoospermic men in Sri Lanka. Asian J Androl *8*, 39-44.

Giachini, C., Laface, I., Guarducci, E., Balercia, G., Forti, G., and Krausz, C. (2008). Partial AZFc deletions and duplications: clinical correlates in the Italian population. Hum Genet *124*, 399-410.

Hucklenbroich, K., Gromoll, J., Heinrich, M., Hohoff, C., Nieschlag, E., and Simoni, M. (2005). Partial deletions in the *AZFc* region of the Y chromosome occur in men with impaired as well as normal spermatogenesis. Hum Reprod *20*, 191-197.

Imken, L., El Houate, B., Chafik, A., Nahili, H., Boulouiz, R., Abidi, O., Chadli, E., Louanjli, N., Elfath, A., Hassar, M.*, et al.* (2007). AZF microdeletions and partial deletions of *AZFc* region on the Y chromosome in Moroccan men. Asian J Androl *9*, 674-678.

Lu, C.C., Zhang, J., Li, Y.C., Xia, Y.K., Zhang, F., Wu, B., Wu, W., Ji, G.X., Gu, A.H., Wang, S.L.*, et al.* (2009). The b2/b3 subdeletion shows higher risk of spermatogenic failure and higher frequency of complete *AZFc* deletion than the gr/gr subdeletion in a Chinese population. Hum Mol Genet *18*, 1122-1130.

Lynch, M., Cram, D.S., Reilly, A., O'Bryan, M.K., Baker, H.W.G., de Kretser, D.M., and McLachlan, R.I. (2005). The Y chromosome gr/gr subdeletion is associated with male infertility. Mol Hum Reprod *11*, 507-512.

Machev, N., Saut, N., Longepied, G., Terriou, P., Navarro, A., Levy, N., Guichaoua, M., Metzler-Guillemain, C., Collignon, P., Frances, A.M.*, et al.* (2004). Sequence family variant loss from the AZFc interval of the human Y chromosome, but not gene copy loss, is strongly associated with male infertility. J Med Genet *41*, 814-825.

Navarro-Costa, P., Pereira, L., Alves, C., Gusmao, L., Proenca, C., Marques-Vidal, P., Rocha, T., Correia, S.C., Jorge, S., Neves, A.*, et al.* (2007). Characterizing partial AZFc deletions of the Y chromosome with amplicon-specific sequence markers. BMC Genomics *8*.

Ravel, C., Chantot-Bastaraud, S., El Houate, B., Rouba, H., Legendre, M., Lorenco, D., Mandelbaum, J., Siffroi, J.P., and McElreavey, K. (2009). Y-chromosome AZFc structural architecture and relationship to male fertility. Fertil Steril *92*, 1924-1933.

Repping, S., Skaletsky, H., Brown, L., van Daalen, S.K.M., Korver, C.M., Pyntikova, T., Kuroda-Kawaguchi, T., de Vries, J.W.A., Oates, R.D., Silber, S.*, et al.* (2003). Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet *35*, 247-251.

Shahid, M., Dhillon, V.S., Khalil, H.S., Sexana, A., and Husain, S.A. (2011). Associations of Y-chromosome subdeletion gr/gr with the prevalence of Y-chromosome haplogroups in infertile patients. Eur J Hum Genet *19*, 23-29.

Stahl, P.J., Mielnik, A., Margreiter, M., Marean, M.B., Schlegel, P.N., and Paduch, D.A. (2011). Diagnosis of the gr/gr Y chromosome microdeletion does not help in the treatment of infertile American men. J Urol *185*, 233-237.

Stouffs, K., Tournaye, H., Van der Elst, J., Haentjens, P., Liebaers, I., and Lissens, W. (2008). Do we need to search for gr/gr deletions in infertile men in a clinical setting? Hum Reprod *23*, 1193-1199.

Visser, L., Westerveld, G.H., Korver, C.M., van Daalen, S.K., Hovingh, S.E., Rozen, S., van der Veen, F., and Repping, S. (2009). Y chromosome gr/gr deletions are a risk factor for low semen quality. Hum Reprod *24*, 2667-2673.

Wu, B., Lu, N.X., Xia, Y.K., Gu, A.H., Lu, C.C., Wang, W., Song, L., Wang, S.L., Shen, H.B., and Wang, X.R. (2007). A frequent Y chromosome b2/b3 subdeletion shows strong association with male infertility in Han-Chinese population. Hum Reprod *22*, 1107-1113.

Zhang, F., Lu, C.C., Li, Z., Xie, P.X., Xia, Y.K., Zhu, X.B., Wu, B., Cai, X.Y., Wang, X.F., Qian, J.*, et al.* (2007). Partial deletions are associated with an increased risk of complete deletion in AZFc: a new insight into the role of partial AZFc deletions in male infertility. J Med Genet *44*, 437-444.

Table S8. Patients with b1/b3 deletions

Patient ID	Spermatogenesis
WHT3453 [a]	Oligospermia [b]
WHT4496	Oligospermia (1.2×10^6 / ml)
WHT4498	Oligospermia ($< 10^6$ / ml)
WHT4755	Oligospermia ($< 5 \times 10^6$ / ml)

[a] Previously reported in S. Repping, et al., "Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection." Nat. Genetics. 35(3):247-251 (2003).

[b] No additional information available.

Supplementary Note

Calculation of mutation rates, relative risk, attributable risk %, and population attributable risk %.

For gr/gr and b1/b3 deletions we estimate the mutation rate using the deterministic model of haploid selection that predicts that, at equilibrium,

$$s = \mu/q$$

where:

s is selection against a genetic variant (in this case the gr/gr or b1/b3 deletion), i.e. the deficit in fitness relative to Y chromosomes without a given deletion,

μ is the rate per generation at which mutations generate the variant, and

q is its prevalence over the entire population (Repping et al., 2003).

Then $\mu = sq$. We have estimates of q from the current study (Table 2, "Prevalence in population"), and we estimate s as follows. First, we require an estimate of F_{SSF}, the proportion of SSF(severe spermatogenic failure) in the unselected population. As discussed in the main text, we estimate this as the fraction of b2/b4 deletions in the population (9 / 20884 from the present study) divided by the fraction of b2/b4 deletions among men with SSF (42/713, ref (Oates et al., 2002)) yielding an estimate of $F_{SSF} = 0.0073$. We estimate confidence intervals by bootstrap resampling (see annotated code at http://jura.wi.mit.edu/page/papers/Rozen_et_al_2012/). Then, let the fitness of the deleted class be $1 - G$, where G is the proportion of SSF in men with the given deletion. (G is presented in Table 2 as a percent as "% of men with deletion who have SSF"). The fitness of the undeleted class is $(1 - P_u)$, where P_u is the proportion of SSF among men without the given deletion. Then, the relative fitness of the deleted class is $(1 - G)/(1 - P_u)$, and the relative loss of fitness,

$$s = [(1 - P_u) - (1 - G)] / (1 - P_u) = (G - P_u)/(1 - P_u).$$

For the gr/gr deletion, this probably represents a lower bound on *s*, as many gr/gr-deleted men without SSF nevertheless have reduced sperm count relative to other men (Visser et al., 2009).

We estimate $G = F_{SSF} * H / q$, where H is the fraction of the given deletion among men with spermatogenic failure. For gr/gr, H is 173/3,935, for b1/b3, 10/3,956, Table 2, "Prevalence among men with SSF"). We estimate $P_u = F_{SSF} (1 - H) / (1 - q)$. Confidence intervals were computed by bootstrapping; please see annotated code at http://jura.wi.mit.edu/page/papers/Rozen_et_al_2012/.

For b2/b4, gr/gr, and b1/b3 deletions we calculate relative risk, attributable risk %, and population attributable risk % as follows:

relative risk = G / P_u.

attributable risk % = $100 (G - P_u) / G$.

population attributable risk % = $100 (F_{SSF} - P_u) / F_{SSF}$.

Confidence intervals were computed by bootstrapping; please see annotated code at http://jura.wi.mit.edu/page/papers/Rozen_et_al_2012/.

References

Oates, R.D., Silber, S., Brown, L.G., and Page, D.C. (2002). Clinical characterization of 42 oligospermic or azoospermic men with microdeletion of the *AZFc* region of the Y chromosome, and of 18 children conceived via ICSI. Hum Reprod *17*, 2813-2824.

Repping, S., Skaletsky, H., Brown, L., van Daalen, S.K.M., Korver, C.M., Pyntikova, T., Kuroda-Kawaguchi, T., de Vries, J.W.A., Oates, R.D., Silber, S., *et al.* (2003). Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet *35*, 247-251.